Open Research Online

The Open University's repository of research publications
and other research outputs

Construction and integration of genome maps of domestic dog chromosomes X and 1

Thesis

How to cite:

Spriggs, Helen Florence (2000). Construction and integration of genome maps of domestic dog chromosomes X and 1. PhD thesis The Open University.

For guidance on citations see FAQs.

© 2000 Helen Florence Spriggs

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.0000ff95

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

Construction and integration of genome maps of domestic dog chromosomes X and 1

Helen Florence Spriggs BSc. (Hons.)

A thesis submitted for the degree of Doctor of Philosophy in fulfilment of the requirements of the Open University

Date of submission: 2nd May 2000

The Animal Health Trust in collaboration with the Sanger Centre

DATE OF SUBMISSION: 25 MAY 2000
DATE OF AWARD: 05 AUGUST 2000

"A dog is a woman's best friend"

ProQuest Number: C805109

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

ProQuest C805109

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

This thesis describes an integrated approach to mapping dog chromosomes X and 1. Small-insert libraries in pBluescript were constructed from flow-sorted chromosomes X and 1 from which, microsatellite-containing clones were identified and sequenced. Database searches identified homologies between X-derived clones and genes previously characterised in other species.

The physical locations of the microsatellite-containing clones were mapped using fluorescence *in situ* hybridisation (FISH), confirming that the clone libraries were highly enriched for their respective chromosomes (91% for chromosome X, 86% for chromosome 1). These libraries, and the mapped clones, represent a valuable resource previously unavailable in dog genomics.

PCR primers were designed from sequences from these libraries, flanking microsatellites for both meiotic linkage and radiation hybrid mapping, and flanking other sequences for radiation hybrid mapping alone. Primers were also designed from other available sequences mapping to dog chromosomes X and 1.

Chromosome X-derived microsatellites were typed on families from the Cornell reference pedigree, adding 17 markers. Chromosome X and 1 markers were typed using the T72 whole genome radiation hybrid panel (WG-RH). A total of 37 chromosome X-derived markers were mapped to eight WG-RH linkage groups and 18 chromosome 1-derived markers were mapped to five WG-RH linkage groups.

Use of a common pool of chromosome-specific markers in three distinct mapping methods facilitates integration, producing unified maps for these chromosomes. The FISH and WG-RH data are complementary, the former giving long-range data and absolute locations, the latter providing local relative ordering. When supplemented with the genetic data, integration with pre-existing maps of chromosome X is possible.

The integrated maps reveal strongly conserved synteny between canine and human X chromosomes. The pseudoautosomal region has been further characterised, the putative or actual locations of nine clinically relevant genes have been suggested, and the canine DMD region better defined.

Acknowledgements

I am indebted to the many friends, family and colleagues who have helped me through the course of this study. It is indisputable that without their collective support I would not have been able to complete the work and hope that in some way I can repay them all in the future. The many and varied forms of assistance are too numerous to mention and I acknowledge all the efforts of the following:

Matthew Binns (supervisor)
Matthew Breen (supervisor)
Julia Carter
Nigel Carter
Paul Dear
Panos Deloukas
Billy Devine
Patricia Fischer
Barbara Gorick
Lisa Hindes
Nigel Holmes (supervisor)
Samantha Johnson
Cordelia Langford
Linda McCarthy
Mark Ross
Penny Rothoff-Rook
Ed Ryder
Tony & Wendy Spriggs
Simon Spriggs
Louise & Glenn Taylor
Rachael Thomas
Bob Wilkins

My thanks to all my colleagues at the Animal Health Trust, past and present who made my time there so enjoyable.

I am extremely grateful for the financial support of the Guide Dogs for the Blind Association who funded my PhD. research and the canine genetics research at the Animal Health Trust.

Contents

Acknowledgements iii
Contents iv
Table of Illustrations vii
List of abbreviations, units and notes ix
1 Introduction 1
1.1 Dog domestication and breeds 1
1.2 Dog characteristics and diseases 2
1.3 The dog as a model organism 3
1.4 Dog karyotype 4
1.5 Genome mapping 6
1.5.1 Markers for genome mapping 7
1.5.1.1 Polymorphic Traits 7
1.5.1.2 Polymorphic DNA sequences 8
1.5.1.3 Defined, monomorphic sequences 11
1.5.1.4 Cloned DNA segments 11
1.5.2 Mapping methods 12
1.5.2.1 Genetic linkage mapping 12
1.5.2.2 Whole genome radiation hybrid mapping 16
1.5.2.3 HAPPY mapping 22
1.5.2.4 Physical mapping 23
1.5.2.5 FISH mapping 23
1.6 Comparative mapping 27
1.7 Dog genome mapping prior to this study 32
1.8 Present study 38
2 Materials and Methods 41
2.1 Flow sorting of canine chromosomes 41
2.2 Small insert libraries - preparation 42
2.3 Small insert libraries - screening for microsatellites and preparation of clone DNA 44
2.4 Small insert libraries - determination of insert size 45
2.5 Small insert clone sequencing 45
2.6 Cosmid clones - screening for microsatellites and preparation of clone DNA 48
2.7 Cosmid clones - end sequencing 48
2.8 Fluorescence *in situ* hybridisation (FISH) 49
2.8.1 Probe preparation 49
2.8.2 Chromosome preparation 50
2.8.3 Hybridisation 50
2.8.4 Probe detection 51
2.9 Sequence analysis and primer design 52
2.10 PCR optimisation 53
2.11 Linkage analysis 54

2.11.1 Assessment of heterozygosity 55
2.11.2 Genotyping of pedigrees 56
2.11.3 Analysis 56
2.12 Radiation hybrid analysis 56
2.12.1 Radiation hybrid typing protocols 57
2.12.1.1 Standard method 57
2.12.1.2 Additional RH typing method from the Sanger Centre 58
2.12.2 Radiation hybrid analysis 59
3 Results 60
3.1 Flow-sorting of chromosomes for library construction 60
3.2 Small insert libraries - preparation 62
3.2.1 Chromosome X 62
3.2.2 Chromosome 1 62
3.3 Small insert libraries - screening for microsatellites 63
3.3.1 Chromosome X 63
3.3.2 Chromosome 1 63
3.4 Small insert clone sequencing 65
3.4.1 Chromosome X 66
3.4.2 Chromosome 1 71
3.5 Cosmid clones - end sequencing 71
3.6 Small insert libraries - determination of insert size 73
3.6.1 Chromosome X 73
3.6.2 Chromosome 1 73
3.7 Sequence analysis 76
3.8 Fluorescence in situ hybridisation (FISH) 78
3.8.1 Chromosome X 78
3.8.2 Chromosome 1 90
3.8.3 Purity of flow-sorting 97
3.9 Linkage analysis 99
3.10 Radiation hybrid analysis 107
3.10.1 Chromosome X 109
3.10.2 Chromosome 1 112
3.11 Integration of FISH, genetic linkage and radiation hybrid data from this study 114
3.11.1 Chromosome X 115
3.11.2 Chromosome 1 117
3.12 Integration of data from this study with existing maps 119
3.12.1 Chromosome X 120
3.12.2 Chromosome 1 124
4 Discussion 126
4.1 Library construction, analysis and microsatellite distribution 126
4.2 Sequence homologies 129

4 Discussion 126
4.1 Library construction, analysis and microsatellite distribution 126
4.2 Sequence homologies 129
4.3 FISH 129
4.4 WG-RH linkage analysis 131
4.5 Genetic Linkage analysis 132
4.6 The value of the integrated mapping approach 133
4.6.1 Chromosome X 134
4.6.2 Chromosome 1 136
4.7 Comparative genomics of chromosome X 138
5 Further work 144
6 Conclusions 145
Appendix 1 147
Appendix 2 154
Appendix 3 160
Appendix 4 175
Appendix 5 178
References 183

Table of Illustrations

Figures

Figure 1 Generation of whole genome radiation hybrids 19
Figure 2 Pedigree structure of the DogMap families 33
Figure 3 Pedigree structure of the Cornell reference families available to the AHT 36
Figure 4 Sequencing strategy for microsatellite-containing pBluescript clones 47
Figure 5 Bivariate flow karyotype of a female dog 61
Figure 6 Screening of small insert clones for microsatellites (96-well format) 64
Figure 7 Screening of small insert clones for microsatellites (384-well format) 64
Figure 8a Electropherogram from clone x75 sequenced with T3 primer 67
Figure 8b Electropherogram from clone x68 sequenced with T7 primer 68
Figure 8c Electropherogram from clone x57 sequenced with DRP $(GT)_{10}A$ 69
Figure 8d Electropherogram from clone x57 with the unique primer designed from DRP sequencing 70
Figure 9 Electropherogram from cosmid H32 with sequenced with T3 primer using BigDye chemistry 72
Figure 10 Insert size analysis of 16 microsatellite-containing plasmid clones 75
Figure 11 FISH of clone x57 to female metaphase spread 82
Figure 12 FISH of clone x71 to male metaphase spread 82
Figure 13 FISH of clone x51 to female metaphase spread 84
Figure 14 Examples of FISH of eight clones showing isolated X or X/Y pairs of chromosomes 84
Figure 15 FISH of clone x28 to female metaphase spread 84
Figure 16 Dual colour FISH of two probes in the DMD region 86
Figure 17 Dual colour FISH two probes to a male metaphase spread 86
Figure 18 Results of FISH of chromosome X-enriched clones 89
Figure 19 FISH of clone 1-32 to metaphase spread 92
Figure 20 FISH of clone 1-33 to metaphase spread 92
Figure 21 FISH of clone 1-18 to metaphase spread 94
Figure 22 Examples of FISH of six clones showing isolated chromosomes 94
Figure 23 FISH of clone 1-46 to metaphase spread 94
Figure 24 Results of FISH of chromosome 1 -enriched clones 96
Figure 25a Genotyping trace from x12 $(CT_4)GT(CT_7)$ with members of the Cornell families 101
Figure 25b Genotyping trace from F9 $(TTTA_n)$ with members of the Cornell families 102
Figure 25c Genotyping trace from x28 $(CTTT_{20})$ with members of the Cornell families 103
Figure 26 Examples of results of optimisation of PCR products for WG-RH mapping 107
Figure 27 Photographs of ethidium bromide-stained agarose gels demonstrating WG-RH typing results 108
Figure 28 Orders of chromosome X markers obtained from two-point analysis using RHMAP 111

Figure 29 Orders of chromosome 1 markers obtained from two-point analysis using RHMAP 113
Figure 30 Integration of the chromosome X mapping data 116
Figure 31 Integration of the chromosome 1 mapping data 118
Figure 32 Taken from Neff *et al.* (1999) showing the Cornell families used to produce the second generation linkage map 121
Figure 33 Linkage mapping results for markers x28 and x51 with data from Neff *et al.* (1999) for sex-averaged linkage groups 11 (top) and 7 (bottom) 123
Figure 34 Linkage results for K338 with markers from the sex-averaged linkage group 1 from Neff *et al.* (1999) 125
Figure 35 Conservation of genes on the X chromosome of eight mammalian species 143

Tables

Table 1 Examples of genome mapping projects and their Internet addresses 28
Table 2 Recombinant and non-recombinant colonies from chromosome X library 62
Table 3 Recombinant and non-recombinant colonies from chromosome 1 library 62
Table 4 Summary of sequencing results from chromosome-enriched libraries 71
Table 5 Summary of BLAST database searching for clones from chromosome enriched libraries 77
Table 6 Cytogenetic position of chromosome X library clones 88
Table 7 Cytogenetic position of chromosome 1 library clones 95
Table 8 Contaminant clones in flow-sorted, small insert libraries 98
Table 9 Polymorphism of markers in the Cornell families. 100
Table 10 Informativeness of markers with respect to Cornell family mapping 105
Table 11 Two-point linkage data from chromosome X markers 106
Table 12 Linkage groups from RH2PT analysis of chromosome-X markers 110
Table 13 Linkage groups from RH2PT analysis of chromosome-1 markers 112
Table 14 Sources of information for genes located on chromosome X in eight species of mammal 141
Table i Primers used in chromosome X mapping 155
Table ii Primers used in chromosome 1 mapping 158
Table iii Genotyping results with chromosome X-derived markers 176
Table iv Typing results of WG-RH panel with chromosome X-derived markers 179
Table v Typing results of WG-RH panel with chromosome 1-derived markers 181

List of abbreviations, units and notes

^{32}P	Phosphorous-32, isotope
A	Adenine
ABC7	ATP-binding cassette-7
AHT	Animal Health Trust
BLAST	Basic local alignment search tool
bp	Base pair
BTK	Bruton's tyrosine kinase
C	Cytosine
CATS	Comparative anchor tagged sequences
CCD	Charge coupled device
cDNA	Complementary DNA
CHM	Choroideraemia
cM	centiMorgan
cR	centiRay
DAPI	4', 6-Diamidino-2-phenylindole
dATP	2'-Deoxyadenosine 5'-triphosphate
DMD	Dystrophin
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DRP	Degenerate repeat primer
dUTP	2'-Deoxyuridine 5'-triphosphate
EDTA	Disodium ethylenediamine tetraacetic acid
FISH	Fluorescence *in situ* hybridisation
FITC	Fluorescein isothiocyanate
F9	Coagulation Factor IX
F8c	Coagulation Factor VIIIc
G	Guanine
g	Gram
GB4	Genebridge 4
HAT	Hypoxanthine, aminopterin & thymidine medium
HPRT	Hypoxanthine guanine phosphoribosyltransferase 1
IPTG	Isopropylthiogalactoside
iu	International units
kb	kilobase
kV	kilovolt
l	Litre
LB	Luria broth
LINE	Long interspersed nucleotide element
Mb	megabase
MER	Medium reiteration sequence

µF	microFarad
µl	microlitre
ml	millilitre
MNK	Menkes syndrome = ATP7A – ATPase, Cu^{2+} transporting, alpha polypeptide
MQ	MilliQ water
ng	Nanogram
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
PDHA1	Pyruvate dehydrogenase (lipoamide) alpha 1
PEG	Polyethylene glycol
PGK1	Phosphoglycerate kinase 1
RF	Recombination fraction
RFLP	Restriction fragment length polymorphism
RH	Radiation hybrid
SINE	Short interspersed nucleotide element
SSC	Trisodium citrate and sodium chloride
STS	Sequence-tagged site
T	Thymine
TAE	Tris-acetate-EDTA buffer
TBE	Tris-borate-EDTA buffer
td	Touchdown
TK	Thymidine kinase
TNG	"The Next Generation"
TOAST	Traced orthologous amplified sequence tags
Tris	Tris[hydroxymethyl]aminomethane
U	Units
UM-STS	Universal mammalian sequence-tagged sites
UV	Ultraviolet light
v/v	volume per volume
w/v	weight per volume
WG-RH	Whole genome radiation hybrid
X-gal	5-bromo-4-chloro-3-indolyl β-Δ-galactopyranoside

Note:

Helen Spriggs was formerly known as Helen Dickens, references prior to 2000 are therefore cited accordingly.

1 Introduction

1.1 Dog domestication and breeds

The domestic dog, *Canis familiaris*, is a member of the family Canidae, in the order Carnivora. There are 34 extant species of canids in 13 genera (Wayne, 1993). They show many variations in size, morphology, behaviour and distribution from the South American jungles, through African savannah, Indian plains and up to the Arctic tundra in North America (Sheldon, 1992, Alderton, 1994). The Canidae also exhibit variation at the subcellular level with chromosome numbers ranging from 36 in the red fox, to 78 in the domestic dog, grey wolf, jackal species, red wolf and wild dogs (Wayne, 1993).

The close association between man and dog extends over many years: it has been estimated that the dog was domesticated more than 100,000 years ago and has been shown to be a direct descendant of the grey wolf (Vilà *et al.* 1997). Over time the dog has been used by man to help catch other animals, to warn of approaching dangers, to herd livestock, to find drugs, contraband, bodies and lost walkers and to assist rescue teams. They have been used to pull sleds, guide blind people, assist deaf and disabled people, to fight other animals and dogs for amusement, catch criminals, protect people and property and, of course, for companionship. The myriad uses to which the dog has readily adapted have resulted in the development of 183 separate breeds of dog recognised by the UK Kennel Club alone (The Kennel Club of Great Britain, 1998). A breed is defined by the Kennel Club as "purebred dogs more or less uniform in size and structure, as produced and maintained by man" (The Kennel Club of Great Britain, 1998). These breeds have been created by generations of selective breeding to achieve a desired set of characteristics. The characteristics for which a breed was originally selected may no longer have a relevant use but the breed is maintained, as it may have other traits which people still prefer. For example, the Bulldog was originally used for bull baiting but its amiable disposition towards humans makes it an ideal family pet in present times. Other breeds, such as the German shepherd dog are now used in roles that exploit their characteristics in a different way: no longer, in this instance, for guarding sheep, but as guard dogs for people and property.

1.2 Dog characteristics and diseases

The intensive inbreeding of dogs to produce the various breeds has resulted in many (approximately 350) inherited diseases (Nicholas *et al.* 1998, Online Mendelian Inheritance in Animals [OMIA] URL: http://www.angis.su.oz.au/Databases/BIRX/omia/omia_form.html). These diseases often occur in a breed or group of breeds due to the founder effect (an individual or line being used frequently as breeding stock). Examples include progressive retinal atrophy (PRA; in English springer spaniels, Dachshunds, Border collies, Toy and Miniature poodles, Golden and Labrador retrievers) and hereditary deafness (in Dalmatians, Bull terriers and Dobermans; Nicholas *et al.* 1998, OMIA).

Of all inherited diseases, about half are incompletely defined; many of these (e.g. hip dysplasia, patent ductus arteriosus) are probably polygenic or involve an interaction of genetic and environmental factors (Patterson *et al.* 1993). However, in many cases the well-established pedigree records makes it easy to define the mode of inheritance of a particular disease. Of the defined diseases, approximately 30% appear to be autosomally transmitted, recessive diseases (Patterson *et al.* 1989), for example, globoid cell leucodystrophy (Krabbe disease), ceroid lipofuscinosis, cystinuria, mucopolysaccharidosis VII, copper toxicosis, progressive retinal atrophy and pseudoachondroplasia (OMIA). A further 8% are autosomal dominant diseases (Patterson *et al.* 1989), for example, congenital lymphedoema, short-limbed dwarfism and nephritis (OMIA). Another 5% are sex-linked (Patterson *et al.* 1989), or have both sex-linked and autosomal forms, for example, progressive retinal atrophy (PRA), tremor (shaking pup syndrome), ectodermal dysplasia, nephritis, severe combined immunodeficiency disease, muscular dystrophy, haemophilia A & B (OMIA).

Genetic testing could provide a means to identify carrier or affected individuals, thereby making it possible to arrange an informed breeding programme in order to reduce the incidence of disease in a particular breed. An example of such a situation is provided by copper toxicosis, an autosomal recessive disease that afflicts Bedlington terriers. This breed has a small population size, and a disease prevalence of 33.9% (Herrtage *et al.* 1987). The disease causes copper accumulation in the liver of the affected animal (Twedt *et al.* 1979) and although it shares some similarities with the human disorder, Wilson's disease, (Tanzi *et al.* 1993, Bull *et al.* 1993), it appears that the gene responsible for copper toxicosis in Bedlingtons is not a canine homologue of the human Wilson's disease genes (Yuzbasiyan-Gurkan *et al.* 1993, van

de Sluis *et al.* 1999). A genetic marker has been identified that is closely linked to the canine disease locus and this makes reliable genetic screening possible (Yuzbasiyan-Gurkan *et al.* 1997; Holmes *et al.* 1998). As this disease is so common in Bedlington terriers, it would not be possible or prudent to immediately stop breeding from all carrier animals: the number of suitable animals would be small and hence only a limited gene pool would be available. This could lead to further problems within the breed if the animals used carried other undesirable traits or diseases. Nevertheless, informed breeding over many generations should make it possible to gradually eliminate this disease from the population.

However, for the majority of dog genetic diseases there are no genetic screens available, hampering efforts to reduce their incidence. Detailed genetic maps can facilitate the identification of disease-linked markers (or the responsible gene itself), making such tests possible. The problems posed by polygenic diseases are more complex, creating an even greater need for genetic maps of the dog genome.

1.3 The dog as a model organism

Dogs offer great potential as model organisms. They have a physiology not that far removed from that of humans, and their long lifespan and large size make them amenable to studies and procedures not practical in many other organisms. These advantages may outweigh the drawbacks, such as their longer generation time and cost of maintenance as compared, for example, to rodents. The dog possesses unique features including the large amount of information in extensive pedigree records for purebred dogs, and the huge range of different temperaments, behaviours and morphologies. This combination of diversity within one species makes them an invaluable tool for comparative genomics.

They are afflicted with a significant number of diseases, both hereditary and acquired, many of which are similar to diseases affecting humans and other mammals. Examples include muscular dystrophy, (Sharp *et al.* 1992, Henthorn *et al.* 1994, Schatzberg *et al.* 1999), mucopolysaccharidosis IIIA (Sanfilippo A; Fischer *et al.* 1998), narcolepsy (Faraco *et al.* 1999, Lin *et al.* 1999), acquired myasthenia gravis (Shelton, 1999) and cancers of many types including, for example prostate cancer (Waters & Bostwick, 1997) and spontaneous squamous cell carcinomas (Gardner, 1996).

The variety of shapes and sizes of dogs that have been produced by selective breeding is remarkable for a single species. For example, dogs conforming to the Kennel Club breed standards range from <12.5cm (Chihuahua) to >87cm (Irish wolfhound) at shoulder height, and from <1kg (Pomeranian) to >90kg (St. Bernard) in weight (The Kennel Club of Great Britain, 1998). Morphological characteristics have been selected to enable the animals to perform a specific function such as the elongated back and short legs of the Dachshund, bred to hunt badgers underground. No other single mammalian species is as diverse in form as the domestic dog and these extremes provide a unique means to study the developmental genetics of morphology.

The generation of most dog breeds has resulted from selecting animals with specific behavioural traits that are of particular use to humans. The availability of these defined characteristics within pedigrees of dogs provides the opportunity to study, for example, behavioural traits such as the herding instinct of Border collies and the Newfoundland's affinity for water (Grandin, 1998). These characteristics clearly have a genetic component and hence provide an unparalleled opportunity to study the genetic constituents of complex behaviour in mammals. Studying well defined, reproducible traits may enable information from dogs to be extrapolated into behavioural traits of humans, although such interpretations need to be handled with great caution, as primate development is much slower than that of other mammals such as the dog. However, dog development is slower than that of other typical model organisms such as the mouse and as such, may offer the opportunity to study traits in an intermediate species.

1.4 Dog karyotype

The dog karyotype has 38 pairs of autosomes and two sex chromosomes, X and Y (Matthey, 1954). It has been estimated to contain 2800Mb of DNA (Langford *et al.* 1996). All the autosomes are acrocentric and vary from 135Mb (chromosome 1) to 39Mb (chromosome 38). The X and Y-chromosomes are metacentric and are 137Mb and 27Mb respectively (Langford *et al.* 1996). The karyotype has taken many attempts to clarify (Selden, *et al.* 1975, Stone *et al.* 1991, Fischer *et al.* 1994, Reimann *et al.* 1996 and Breen *et al.* 1999b). It has proved difficult for many of the researchers to consistently produce elongated metaphases from which the subtle differences between the smaller chromosomes may be observed. Probably because of this,

there have been inconsistencies between groups as to the assignment of chromosome names. There have been efforts to produce a standardised karyotype (Switonski *et al.* 1996) and although the first 21 autosomes and the sex chromosomes have been standardised, the smaller autosomes, 22 – 38 have proved difficult. One group have reported an extended nomenclature with all of the chromosomes identified using a combination of GTG and RBG-banding (Reimann *et al.* 1996). More recently, the International Standardisation Committee agreed on the assignments of autosomes 22 – 38, thus providing dog genome researchers with a definitive dog karyotype based on classically banded chromosomes and the use of flow-sorted chromosome paints (Breen *et al.* 1998, Breen *et al.* 1999b).

The X and Y chromosomes are relatively easy to identify: they are both metacentric, and are the largest (137Mb) and smallest (27Mb) chromosomes, respectively (Langford *et al* 1996). In humans, the sex chromosomes are of similar size (approximately 164Mb and 28Mb long respectively; Schuler *et al.* 1996). In March 2000, the Genome Database, (GDB, http://www.gdb.org/) described 566 human genes and 29 pseudogenes that have been mapped to human chromosome X, and 48 genes and nine pseudogenes to chromosome Y. In addition, this database describes 217 genes for genetic disorders known to be located on chromosome X and ten on the Y chromosome. The database also demonstrates that the density of disease-related loci on human chromosome X is more than twice that of any of the autosomes. A similar density of genes and genetic disorders would be expected on the dog sex chromosomes.

In all mammals, one copy of the X-chromosome is subject to X-inactivation in female (XX) cells, to maintain dosage compensation (Lyon, 1961 in Carrel *et al.* 1999). The majority of the genes in one of the X chromosomes are inactive, to ensure equal levels of expression with those seen in male (XY) cells. The exceptions to this rule are genes found in the pseudoautosomal regions (PARs) of the X chromosome, which have counterparts on the smaller Y chromosome and are typically expressed from both copies of X in female cells, or from both X and Y in male cells (Goodfellow *et al.* 1983, Graves *et al.* 1998a). Recombination also takes place between the PARs of X and the corresponding regions of Y in male cells, much as it does between the paired X chromosomes in female cells, or between paired autosomes in cells of either sex. In man, the major PAR (PAR1, Burgoyne, 1982) is located at Xp, measures 2.6Mb, and shows recombination with Yp (Graves *et al.* 1998b). Recombination involving PAR1 in female cells is about ten-fold lower than in males, presumably because recombination between other regions of the paired X-chromosomes in females suppresses recombination in

PAR1 through positive interference (Page *et al.* 1987; Hunt et al, 1992, Schmitt *et al.* 1994). A second human PAR (PAR2) is located at the Xq telomeric region (recombining with Yq; Freije *et al.* 1992); it is far smaller (~320kb) than PAR1and undergoes less frequent recombination (Kvaløy, *et al.* 1994).

In humans, eight genes have been mapped to PAR1 (adenine nucleotide translocator 3, [ANT3], acetylserotonin methyltransferase, [ASMT], colony-stimulating factor 2 receptor, alpha, [CSF2RA], interleukin-3 receptor, alpha, [IL3RA], surface antigen MIC2, [MIC2], pseudoautosomal GTP-binding protein-like, [PGPL], short stature homeo box, [SHOX] and pseudoautosomal gene XE7, [XE7], as has the 5' region of a ninth gene, the XGa blood group system, [XGA] (Graves *et al.* 1998b). Two genes, interleukin-9 receptor gene, (IL9R) and synaptobrevin-like gene, (SYBL1) have been mapped to PAR2 (Freije *et al.* 1992, D'Esposito *et al.* 1996, Kermouni *et al.* 1995). For functional reasons, homologues of many or most of these genes might be expected in the canine PAR. The segregation pattern of the loci indicated that X/Y recombination results from a single obligatory meiotic crossover in the PARs. PAR1 is characterised by a high recombination activity that lends itself to extensive mapping. In addition to PAR1 and PAR2, there is a much larger region of approximately 4Mb known to share X/Y homology in man but it is not believed to pair during meiosis (Mumm *et al.* 1997). It is also believed to be unique to humans having been shown to be located only on the X chromosome (i.e. not on the Y chromosome) of chimpanzees and lower vertebrates (Page *et al.* 1984, 1987, Geldwerth *et al.* 1985, and Koenig *et al.* 1985).

1.5 Genome mapping

The study of the complex interactions of genes in health and disease requires detailed knowledge of genome structure and organisation. A genome map gives the location of many unique reference points (markers) in the genome. Depending on the type of markers used and the method by which they are mapped, the map can be used to localise and identify genes, to compare the genome organisation of different species, or as a basis for genome sequencing. Typically, a low-resolution genetic map (giving the positions of polymorphic traits or sequences, deduced from their patterns of segregation) will allow the approximate location of the gene responsible for a disease or trait to be determined; physical maps can then be used to analyse

the region of interest at the molecular level. The following sections introduce the various types of marker that can be placed on genome maps, and the methods by which they can be mapped.

1.5.1 Markers for genome mapping

Genomes contain sequences with known functions such as protein-coding regions, as well as those with less well characterised roles such as repetitive elements, or areas with no known function, so called "junk DNA" (Csink & Henikoff, 1998). A marker on a genome map may be a unique, defined fragment of DNA sequence from any of these regions. However, uncharacterised DNA fragments (for example, cloned fragments of unique but unknown sequence) may also serve as markers. A phenotypic trait (such as a disease) may serve as a marker, the phenotype reflecting the underlying (but unknown) DNA sequence. The major categories of markers are outlined in the following sections.

1.5.1.1 Polymorphic Traits

Polymorphic traits were the first genetic factors to be mapped, before the advent of molecular genetics. Sturtevant (1913) first noted that certain mutations in *Drosophila melanogaster* were inherited in a sex-linked manner, and reasoned that their pattern of inheritance (segregation) reflected the physical arrangement of underlying (but undetectable) genetic factors. This observation led to the development of genetic linkage mapping (see section 1.5.2.1). Any phenotypic variation which is inherited in a simple Mendelian manner can serve as a marker in genetic mapping: the trait acts as a proxy for the underlying gene, and many genes (particularly those responsible for diseases or other abnormalities) are initially defined and mapped in this way. In man, over 10,000 diseases and traits showing simple Mendelian inheritance are catalogued in the database OMIM (Online Mendelian Inheritance in Man; http://www.ncbi.nlm.nih.gov/omim/). A similar database for animal diseases has been developed although this currently only includes literature citations (OMIA).

Many variable traits with a genetic basis do not show simple Mendelian inheritance. Such traits (for example, body weight, intelligence or growth rate) are often continuously variable, and are the result of the interaction of many genes, or of a combination of genetic and environmental factors. These complex traits (often referred to as quantitative traits) cannot be

used as simple genetic markers, but are the subject of intensive study in the context of human disease and in agriculture.

1.5.1.2 Polymorphic DNA sequences

Although early genetic maps of *Drosophila*, yeast, mouse and man employed polymorphic traits as markers, there are a limited number of such traits available in most species. Advances in molecular genetics, however, made it possible to identify a far greater number of DNA sequence polymorphisms that had no observable phenotypic consequence. Such polymorphic DNA sequences have made it possible to make more detailed genetic maps. The first of these polymorphic markers were restriction fragment polymorphisms (RFLPs, Botstein *et al.* 1980). RFLP mapping uses restriction digestion of the DNA of interest to examine differences in distribution of restriction sites. Originally, these markers were mapped using Southern blots from restriction digests of the test DNA, and subsequent hybridisation with radioactive probes (Botstein *et al.* 1980). The advent of the polymerase chain reaction (PCR) simplified the mapping process. The sequence is amplified by the PCR and the product tested for cleavage with the appropriate restriction enzyme. RFLPs typically have only two alleles (the restriction site is either present or absent) and this limits their informativeness for genome mapping.

RFLPs have largely been supplanted as polymorphic markers by variable number tandem repeat elements (VNTRs) such as mini- or microsatellites. Such elements consist of a unique sequence that contains within it a tandemly repeated motif. The number of repetitions varies from allele to allele, resulting in length polymorphism in PCR products produced by primers designed in the unique flanking sequence. Microsatellites (containing tandemly repeated motifs of 2 - 6 bases) are especially useful for mapping since they are widely dispersed in the genome, are easily detected and analysed by the PCR, and often show high degrees of polymorphism (Tautz, 1989). There is one microsatellite repeat, on average, every 10kb in the human genome (Tautz, 1989). Of these, $(CA)_n$ repeats are the most common (Love *et al.* 1990), and have been estimated to occur about every 30kb in man (Stallings *et al.* 1991, Beckman & Weber, 1992), every 18kb in mice (Stallings *et al.* 1991) and every 42kb in dog (Rothuizen *et al.* 1994). Although mostly found in intergenic DNA or in introns of genes, some microsatellites (so far, only trinucleotides) have been recorded in the coding sequence of genes, and have been implicated in the aetiology of some diseases. Triplet expansion has been

shown to be involved in certain diseases, for example, $(CAG)_n$ in Huntingdon's disease (Huntington's Disease Collaborative Research Group, 1993), spino-cerebellar ataxia 1 (Orr *et al.* 1993) and spinal and bulbar muscular atrophy (La Spada *et al.* 1991).

The length polymorphism of microsatellites has been demonstrated (Tautz, 1989, Weber & May, 1989) and may be caused by replication slippage within the repetitive stretches (Tautz & Renz, 1984). The stability of microsatellites from generation to generation has been observed by their Mendelian inheritance, which has enabled their use as markers for linkage mapping (Tautz, 1989, Weber & May, 1989). The polymorphism of a repeat is related to its length: repeats of less than ten units tend to show little polymorphism, those with lengths of between eleven and fifteen have variable polymorphic properties whilst larger repeats show consistently moderate to high polymorphism (Weber, 1990). Three categories of microsatellite repeat have been described (Weber, 1990):

- Perfect repeat sequences are defined as alternating, tandem repeats without interruption and without adjacent repeats from another sequence, e.g.: $(CCG)_{12}$, $(CA)_{15}$ or $(GAAA)_{15}$.
- Imperfect repeats are two or more runs of uninterrupted repeats separated by less than three consecutive nonrepeat bases. Terminal runs of uninterrupted repeats (outside of nonrepeat bases) must be at least three full repeats in length. The sequence: $(CA)_{22}GACACAC$ would therefore be classified as a perfect repeat sequence with 22 repeats. The sequence $(AC)_{12}GT(AC)_{10}$ would be designated as a single imperfect repeat sequence, but the sequence, $(GT)_{15}TACG(GT)_6$ would be considered as two separate perfect repeats.
- Compound repeats are defined as runs of repeats, separated by no more than three consecutive nonrepeat bases from a run of ≥5 uninterrupted dinucleotide or longer repeat length repeats of a sequence other than the $(CA)_n$ form, or from ≥10 uninterrupted mononucleotides. Compound repeats are subclassified as perfect or imperfect depending on the status of the $(CA)_n$ block, for example: $TCTCT(AC)_{16}(TC)_{10}AGC$ is a perfect compound repeat, whereas $CCAT(CT)_{12}GTTT(TC)_{11}T(CA)_{14}A(AC)_6$ CAC is an imperfect compound repeat

Tetranucleotide repeats are highly polymorphic and less prone, than shorter repeats, to the formation of artefact bands in the PCR (He *et al.* 1996). They have been estimated to occur every 130kb in dog (Rothuizen *et al.* 1994, Francisco *et al.* 1996).

A wide variety of other sequence polymorphisms can also be used as markers in genetic mapping. For example, Miller (1994) detected and exploited length polymorphisms between repetitive elements such as short interspersed elements (SINEs) or long interspersed elements (LINEs) and nearby unique sequences. The positions of such elements may be conserved between species (for example, a SINE element is found in the 3' untranslated region of both sheep and goat ZXF genes; Xiao *et al* 1998), making associated polymorphisms particularly valuable.

Two types of loci have been described that are used in making maps (O'Brien *et al.* 1993): conserved coding loci, referred to as type I markers, and highly variable sequences, such as microsatellites, referred to as type II markers. These two types of loci have distinct values for physical or genetic mapping. Coding sequences are valuable for comparative mapping because they are conserved between species, orders or even classes and homologies can be established. Cloned coding genes may be used for *in situ* hybridisation within or across species. Microsatellites (type II loci) are common in vertebrates but are usually located in introns or untranslated regions. The variation in these type II loci makes them valuable for constructing genetic maps. However, since most type II loci are species specific they have limited use in comparative mapping, although in closely related species such as cattle and sheep (Kemp, *et al.* 1995, de Gortari *et al.* 1998, Nijman *et al.* 1998) or rat and mouse (Kondo *et al.* 1993) it has been shown to be possible to use these markers between species. The use of type I and type II markers together to generate both physical and genetic mapping data is extremely beneficial to the genome mapping efforts and the integration of the maps via common markers makes them very valuable (Lyons *et al.* 1997).

Single nucleotide polymorphisms (SNPs) are an abundant form of genome variation, they are distinguished from rare variations by a requirement for the least abundant allele to have a frequency of 1% or more. In the human genome SNPs are estimated to occur approximately once per kilobase (Brookes, 1999). They are of increasing interest because of their potential to identify the genes underlying polygenic diseases (Collins *et al.* 1997). SNPs can have a maximum of four alleles (and typically have only two), and are therefore less informative than microsatellites, but they can be typed in high-throughput systems, such as multiplex PCR and oligonucleotide microarray technology.

The main drawback of the majority of polymorphic markers (with the exception of RFLPs detected by Southern hybridisation) is the need to know the sequence surrounding the

polymorphism. Other types of polymorphic marker have been developed to overcome this, for example, amplified fragment length polymorphisms (AFLPs; Caetano-Anolles *et al.* 1991, Kuiper, 1998) and random amplified polymorphic DNA markers (RAPDs; Williams *et al.* 1990). These markers use the detection of length variations in arbitrarily amplified DNA sequences: each amplified fragment of DNA may be considered as a 'locus' (or an amplifying site) and follows the principles of Mendelian inheritance.

1.5.1.3 Defined, monomorphic sequences

In principle, any unique, known region of sequence can be used as a marker, although such markers cannot be mapped by genetic means as they do not have multiple alleles which can be used to show segregation. Such markers are referred to as sequence tagged sites (STS). Primers are designed to flank the region and used in the PCR to detect the presence of the STS in the target sequence, for example from genomic DNA, somatic cell hybrid DNA or radiation hybrid DNA. If the sequence is obtained from a known gene sequence, the marker is called an expressed sequence tag (EST). Mapping ESTs allows the positions of the associated genes to be determined. STS and EST markers are generally not polymorphic and are used in physical mapping techniques.

1.5.1.4 Cloned DNA segments

The development of a number of techniques for ligating, propagating and selecting DNA fragments has enabled the cloning of defined, isolated genomic DNA fragments. The clones are typically produced by either random mechanical shearing, or restriction digestion of genomic DNA. This material is then ligated into a suitable vector, the choice of which will depend on the use to which the cloned DNA will be put. For genome mapping, it is often preferable to have relatively large cloned fragments so that the numbers that have to be analysed are reduced. Vector systems that can contain large fragments include cosmids (plasmids which can be packaged and transfected like phage lambda systems) such as pWE15, which can contain fragments of up to 44kb (Stratagene), P1 bacteriophage artificial chromosomes (PAC) vectors, for example, pCYPAC2, that can hold up to 150kb (Stratagene), and bacterial artificial chromosomes (BAC) vectors, for example, pBACe3.6 which can hold inserts of up to 300kb (Stratagene). Yeast artificial chromosomes (YAC), for example, pYAC4, have the largest insert capacities (0.2 - 2Mb), but are more difficult to manipulate than bacterial clones and are prone

to a variety of cloning artefacts that can complicate mapping, for example clone chimaerism (Monaco & Larin, 1994).

Such cloned fragments can be located in the genome (and hence serve as markers) using a variety of techniques, without requiring extensive sequence information or other characterisation of the cloned fragment. Examples of such techniques (described in section 1.5.2) include FISH and contig assembly.

Flow sorting can be employed to isolate pools of DNA enriched for chromosomes and has been used in a number of species including man (e.g. Nižetić *et al.* 1991, Longmire *et al.* 1993), pig (Miller *et al.* 1992) and mouse (Weier *et al.* 1994). Flow-sorted libraries have been constructed and used for a variety of mapping projects, for example, human chromosome X and 21-specific libraries have been produced (Nižetić *et al.* 1991) and used to establish ordered libraries of overlapping cosmid clones. The practicability of using this technique in another mammalian species has been demonstrated for example, in the pig (Miller *et al.* 1992), where a small insert, flow-sorted chromosome 1-specific library was produced and used to identify and isolate dinucleotide repeats specific for the chromosome (Anderson-Dear & Miller, 1994). Similarly, Grimm *et al.* (1997) successfully flow-sorted pig chromosome 6, cloned it into pBluescript to produce a small-insert library and used it to screen for pig chromosome 6-specific microsatellite markers.

1.5.2 Mapping methods

There are a number of mapping methods available with different ranges and resolutions, which may be used independently or in combination to form an integrated map. For example, fluorescence *in situ* hybridisation gives the absolute locations of cloned probes on the chromosomes, albeit at low (2 – 3Mb) resolution. In contrast, whole genome radiation hybrid mapping can provide, at high (50kb) resolution, the relative positions of STS markers, but fails to give the absolute chromosomal locations.

1.5.2.1 Genetic linkage mapping

Genetic mapping uses the estimation of frequency of meiotic recombination between two loci; this frequency increases as the distance between two loci increases. Loci on different chromosomes will undergo independent assortment at meiosis according to Mendel's second

law. If they are close together on the same chromosome, they are likely to remain associated rather than being separated by a chiasma, and hence will appear linked and co-segregate in most meioses. The closer two genes lie together; the more tightly they will be linked. Recombination occurs frequently and at least one crossover must occur during meiosis per chromosome. The occurrence of one chiasma reduces the probability of another forming close by; this effect is called positive interference (Page *et al.* 1987; Hunt et al, 1992, Schmitt *et al.* 1994).

Sturtevant, (1913) exploited the frequency of recombination to generate the first genetic map. A number of *D. melanogaster* mutants were studied where the causative genes were known to be located on the X-chromosomes because of the observed sex linkage. When recombination studies were performed on these mutants, the recombination frequencies observed between different pairs of genes were found to depend upon precisely which pairs of genes were being studied. These data were used to show the relative positions of the genes along the X chromosome and to show that the map must be linear in common with the observation of the thread-like form of chromosomes under the microscope.

Genetic mapping provides the only means by which loci responsible for specific phenotypes (such as diseases) may be mapped, in the absence of any molecular knowledge of the underlying genes. As such, it remains an invaluable tool in genomics. Nevertheless, only polymorphic loci (which permit the segregation pattern to be observed) can be mapped, and the number of phenotypically observable polymorphic loci in any one species is usually small, limiting the density of maps that can be made. Therefore, dense genetic linkage maps are normally made using markers based on polymorphic DNA sequences (section 1.5.1.2), which have no observable phenotype. The recombination fraction between two loci is defined as the proportion of recombinants (ranging from 0 for tightly linked loci to 0.5 for loci which are unlinked and therefore segregate independently). It is usually expressed in centiMorgans (cM), where 1cM corresponds to a recombination fraction of 1%. (The genetic distance between unlinked loci is therefore 50cM). The conversion of genetic distances into physical distances requires a model (the mapping function) which takes into account the distribution of crossovers along the chromosomes, the likelihood of more than one crossover between two loci, and the effect which one crossover event has on the likelihood of other crossovers nearby (such as positive interference, above). This mapping function typically differs between males and females, and therefore genetic map distances are normally given as male-specific, female-specific and/or

sex-averaged. (An exception is found in cattle, where little difference in length between the male and female maps has been observed [Kappes *et al.* 1997].) Chiasmata counts observed in human male meioses show an average of 49 crossovers per cell (Morton *et al.* 1982). Each crossover gives 50% recombinants so the chiasmata count of 49 indicates a male genetic map length of 2450cM (i.e. 50% recombination x 49 crossovers). The best estimate from linkage mapping, adding together all the chromosome lengths, is 2730cM (Broman *et al.* 1998). Chiasmata are more frequent in female meioses, in part because the two X-chromosomes in females can pair along their whole length whereas the X and Y-chromosomes in males can only recombine at the relatively small pseudoautosomal regions. The total human female map length determined by Broman *et al.* (1998) was 4435cM. Thus over the 3154Mb of the human genome (Schuler *et al.* 1996), one male cM averages corresponds on average to 1.15Mb and one female cM averages to 0.71Mb, the sex-averaged figure being 0.9Mb. However, the actual correspondence varies widely for different chromosomal regions, because the distribution of chiasmata is not uniform, on both a large and small scale. The extreme deviation is shown by the pseudoautosomal region at the tips of the p arms of the human X- and Y-chromosomes (PAR1). Males have an obligatory crossover within this region, which measures 2.6Mb in humans, so that it is determined to be 50cM long, whereas in females the same region has a genetic length of 7cM. Therefore, for this region in human males, 1Mb = 19cM, whereas in females 1Mb = 2.7cM. The lack of recombination between mammalian X and Y chromosomes (other than in the pseudoautosomal regions) makes it impossible to produce genetic maps of the Y chromosome, or to measure genetic distances on the X chromosome in males. Typical recombination of X-chromosomes occurs in females and therefore genetic mapping of the X-chromosome is possible in female meioses.

In most mammals, it is not possible to observe all the recombination events; therefore, a statistical calculation is used to determine the likelihood of linkage between markers, based on the available data (Terwilliger & Ott, 1994). In the maximum likelihood approach, a range of candidate recombination fractions (θ) from 0 to 0.5 are tested, to find that which has the highest likelihood, L_θ, of producing the observed segregation data. This value of θ is then considered the best estimate of the true recombination fraction. The confidence in declaring the two loci to be linked is given by the ratio of L_θ to $L_{0.5}$ - the likelihood of obtaining the observed results on the hypothesis that the loci are unlinked. This confidence is normally expressed as a logarithm (base 10), termed the lod score between the loci. Hence:

$$Z_\theta = \log_{10} \frac{L_\theta}{L_{0.5}} \qquad \text{(Equation 1)}$$

Where Z_θ is the lod score

A lod score of 3 at $\theta = 0.2$ means that the observed pattern of segregation is 1000-fold (10^3) more likely to have arisen due to linkage between the markers (at a distance of $\theta = 0.2$) than by chance alone between unlinked markers ($\theta = 0.5$). A lod score of >3 is generally accepted as being the minimum value to declare linkage between markers on autosomes, giving acceptable confidence when weighed against the prior odds against linkage between autosomal markers (Ott, J. in Human Genetic Disease: a practical approach, Davies, K. E (Ed) 1986, pp19 - 32). For markers on the X chromosome, lod scores of 2.0 are acceptable for linkage because knowledge of the chromosomal location of both markers reduces the prior odds (Ott, 1991). The informativeness of lod scores is related to the number of informative meioses; therefore, where lod scores are less than 3.0, typing more individuals may increase the lod score and provide evidence of linkage (Terwilliger & Ott, 1994).

Genetic linkage mapping requires reference pedigrees of families, for example, the Centre d'Étude du Polymorphisme Humaine (CEPH) families in man, the European Collaborative Interspecific Backcross (EUCIB) in mouse, or Illinois Reference/Resource Families (IRRF) and International Bovine Reference Panel (IBRP) in cattle. Some domesticated mammals have extensive pedigree records that have been kept for many generations. For example, the first general studbook for thoroughbred horses was published by Weatherby's in 1791 and the Kennel Club studbook for dogs was first published in 1873. Such pedigree data, coupled with knowledge of the disease status of animals and where possible, DNA samples permit the inheritance of a phenotype or disease to be followed and mapped, (unless the disease has arisen from a recent, spontaneous mutation, in which case the previous generations will not have the disease). Loci associated with a disease can similarly be followed, which enables information about a disease to be gleaned without actually knowing the molecular characteristics for the disease.

In addition to being polymorphic, it can also be seen that the markers must be heterozygous in the parents if segregation is to be observed in the offspring. Individuals that meet this requirement are said to be informative for the locus. Maximum information on the linkage between two loci (A and B) can be obtained only when the linkage phase - the physical

association of marker alleles with each of the homologous chromosomes in the informative parent - is known (i.e., whether the heterozygous parent is AB/ab or Ab/aB). The lod score calculation for phase-unknown parents is less informative, as it has to account for either of the possible parental phases and the overall result is obtained by summing each lod score weighted by its prior probability. A minimum of three generations of animals is normally required to determine the linkage phase and hence two-generation pedigrees have limited power in genetic linkage mapping.

The calculations involved in estimating genetic distances from pedigree data are complex. Therefore, a number of computer programs have been developed for linkage analysis both of human pedigree data and of the more complex and inbred animal pedigrees commonly available. Those most frequently used are part of a package called Linkage (Lathrop *et al.* 1984).

Once pairwise distance estimates have been calculated between many loci, they must be integrated to produce a map giving the order and distance between all loci. Ordering markers on genetic linkage maps is a huge task, for which a rigorous formal solution is not possible for significant numbers of markers. One common approach is to construct an initial framework map based on a small number of loci, and then place additional markers relative to these by multipoint linkage mapping; the use of other methods such as physical assignment of markers assists the problem considerably. CRI-MAP (Lander & Green, 1987) is a multipoint linkage analysis programme originally developed to map RFLPs on the CEPH families; however, it can be used to analyse microsatellite markers on other pedigrees (Green, *et al.* 1990).

1.5.2.2 Whole genome radiation hybrid mapping

Genetic linkage mapping suffers from a number of drawbacks: polymorphic markers must be available, it requires informative pedigrees of animals and the resolution of the resulting maps is limited (typically to around 1cM ≅ 1Mb) by the frequency of meiotic recombination. The need for higher resolution mapping led to the development of whole-genome radiation hybrid panels (WG-RH) by Goss & Harris (1975). In RH mapping, donor cells (containing the genome to be mapped) are irradiated to randomly break their chromosomes, and then rescued by fusion with unirradiated host cells of a different species to form hybrids. In the early experiments, the hybrids were analysed for their production of donor proteins (reflecting the presence of donor

chromosome fragments). The hybrids were found to lose donor fragments during culture, eventually retaining only a small proportion of the donor genome. It was reasoned that the pattern of retention of donor genes (identified by their protein products) would reflect their organisation along the chromosomes: adjacent donor genes would tend to be jointly retained (or jointly lost) in the hybrids. The original form of this method was not generally utilised, with the exception of some human chromosome X mapping (for example: Willard *et al.* 1985). When the method was revisited (using molecular genetic techniques rather than protein analysis to identify donor genes), it was mainly applied to chromosome-specific panels where the donor cell was itself a hybrid containing a single human chromosome. Walter *et al.* (1994) returned to the methods of Goss & Harris to produce a whole-genome radiation hybrid (WG-RH) panel using diploid human fibroblasts as the donor and rodent cells as the host. The method for creating a WG-RH panel is outlined in figure 1. The donor cell line is exposed to a given dose of X irradiation; the irradiated donor cells are fused with equal numbers of the recipient rodent cells using polyethylene glycol (PEG). The donor cell line contains a selectable marker, such as thymidine kinase (TK), for which the recipient cell line is deficient. If thymidine kinase is used as the marker, the cells are grown in medium containing hypoxanthine, aminopterin and thymidine (HAT). Cells deficient in thymidine kinase (TK-) cannot survive, as they cannot metabolise the substrates in HAT. Therefore, unfused recipient (TK-) cells are counterselected, whilst unfused donor cells are killed by the irradiation. The surviving hybrids may contain the donor chromosomal material attached to the recipient's chromosomes or the donor fragments may be reconstituted as "new" donor chromosomes (Walter *et al.* 1994). The hybrid colonies are picked into plates and expanded by culture for DNA extraction. Confirmation of the donor genome content of the hybrids can be carried out using a variety of methods. Standard cytogenetic analysis can reveal the crude composition of each hybrid (McCarthy *et al.* 1997). Metaphase spreads of the donor, recipient and hybrid cell lines may be prepared and fluorescence *in situ* hybridisation (FISH) performed using labelled total genomic DNA from the donor species. The content of human cell hybrids can also be identified by *Alu*-PCR, which selectively amplifies human sequences by priming on human-specific *Alu* repeat elements. *Alu*-PCR products are specifically amplified from *Alu* repetitive elements in human genomic DNA, which can be labelled and FISH used to examine the contents of the hybrid cell lines (Nelson, *et al.* 1989, Bicknell *et al.* 1991). Once the panel of hybrids (typically 80 - 90 hybrid cell lines) has been established, DNA from each hybrid may be screened by PCR to determine which donor markers

it retains. If two markers are closely linked in the donor genome, then the random irradiation-induced breakage will seldom break them apart; therefore, they will tend to be co-retained in hybrid cells, and will be seen to "co-segregate". Conversely, donor markers that are further apart will normally be broken apart during irradiation, to lie on independent fragments. They will therefore show no particular tendency to be retained together amongst the hybrids, and will not co-segregate. This forms the basis for estimating distances between markers. The average size of the chromosomal fragments is inversely proportional to the dosage of radiation (Boehnke *et al.* 1991), which therefore determines the resolution of the hybrid panel. High doses of radiation produce small fragments, allowing high-resolution maps to be made. Lower doses give larger fragments, producing poorer map resolution but allowing more widely spaced markers to be mapped.

In the human WG-RH panel, Genebridge 4 (GB4), produced with a radiation dose of 3000rads (Gyapay *et al.* 1996), some of the human fragments were found to be integrated into the host genome, whilst others were maintained as reconstituted human chromosomes. These may be the result of human centromeres recruiting human telomeric sequences to form "new" human chromosomes (Walter *et al.* 1994). More recently, human panels have been produced with higher radiation doses to enable higher resolution mapping. These include the Stanford G3 panel (10,000 rads; Stewart *et al.* 1997) and TNG panel (50,000 rads; Lunetta *et al.* 1996). The dog panel, T72 (Research Genetics) showed a pattern of integration into the host chromosomes similar to the human GB4 panel, as well as containing reconstituted dog chromosomes (L. McCarthy personal communication). However, in the mouse WG-RH panel, T31, the majority of the mouse chromosomes were not integrated into the host hamster chromosomes but were maintained as independent recombinant mouse chromosomes in the hybrids (McCarthy *et al.* 1997).

Radiation hybrid (RH) mapping can be used, in principle, to map any sequence-tagged site (STS) regardless of polymorphism. The STS should meet certain criteria: it should be reliably amplified by PCR, be present in a single copy in the donor genome and should not be closely homologous to a sequence in the host genome (which would co- amplify during marker screening).

In many ways, RH mapping is similar to linkage mapping and the statistical methods used to interpret the data are analogous. The initial step is to calculate two-point lod scores between pairwise combinations of markers, based on their co-segregation frequencies amongst

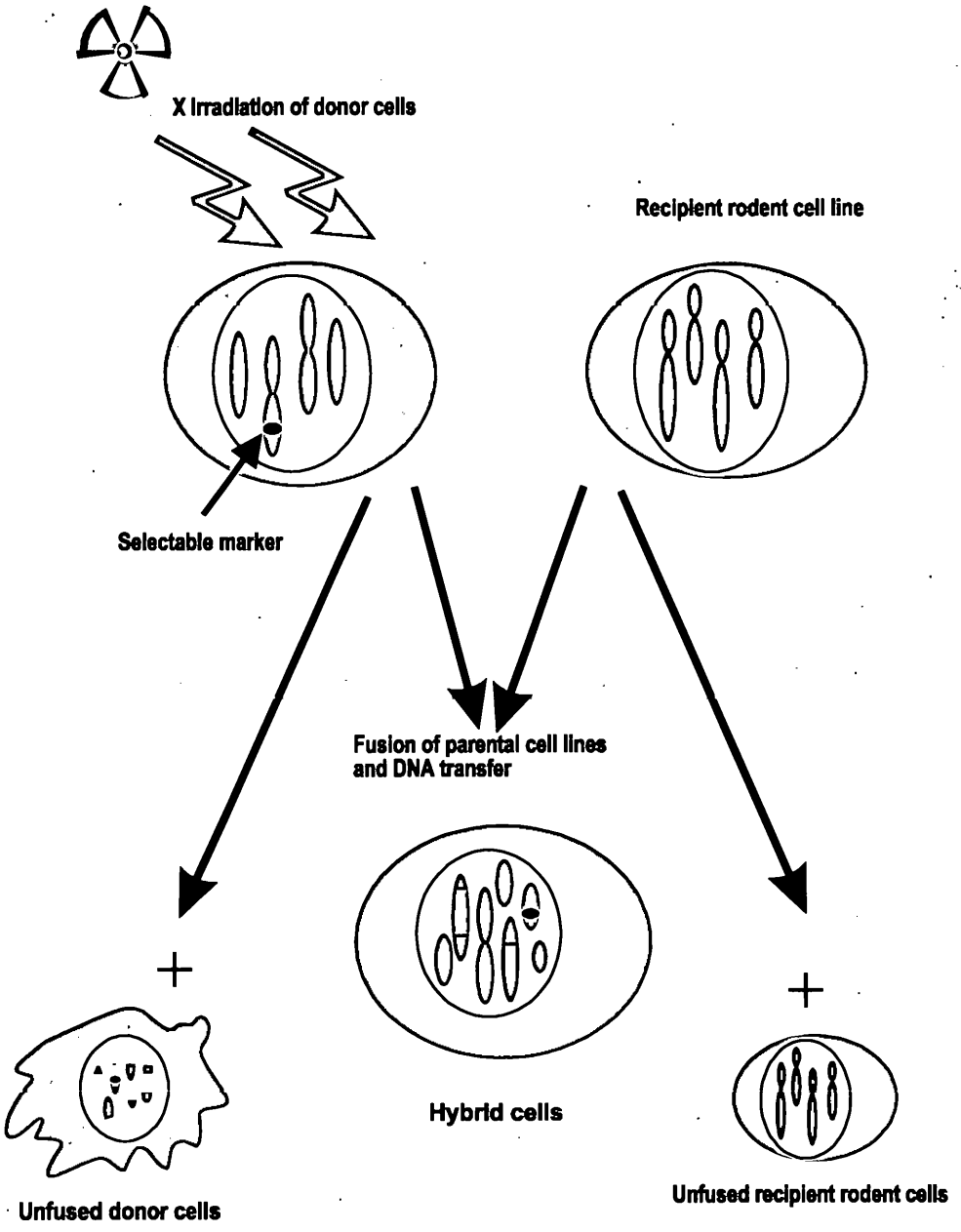


Figure 1 Generation of whole genome radiation hybrids.
Donor cells containing a selectable marker are irradiated with a given dose of X-irradiation and fused with the recipient, rodent cell line. The non-recombinant rodent cells are selected against and the donor cells die leaving only the hybrid cells containing fragments from the donor cells. The donor chromosomes may be present as independent donor chromosomes formed from the fusion of two donor fragments, or may be fused with the recipient chromosomes. The DNA associated with the selectable marker will be present at high frequency.

the hybrids. The associated distance estimate (θ) is analogous to that in linkage mapping, but can have values from 0 to 1 (reflecting the probability of radiation-induced breakage between the two markers). A number of analysis packages are available, for example, RHMAP (Lange *et al.* 1995), RHMAPPER (Stein, L. 1996, Slonim *et al.* 1997), RADMAP (Matise *et al.* 1994) and SAMAPPER (Stewart *et al.* 1997).

The original software for RHMAP (Boehnke *et al.* 1991) was developed for single chromosome RH mapping but was adapted (Lange *et al.* 1995) to handle whole-genome and polyploid data. This software first generates two-point linkage data, which allow estimation of the distance between two markers in the data set and identify groups of linked markers. Based on these results, markers are sorted into groups that show linkage to one another at or above a specified threshold lod. The two-point data can then be further analysed to determine the order and spacing of the markers. This is initially carried out to identify a preliminary order for some or all of the markers and can be done by calculating the minimum number of obligate chromosome breaks for the locus order, since two loci lying close together on the chromosome should have fewer breaks between them than two lying far apart (Boehnke *et al.* 1991). Thus, the best locus order is that requiring the fewest obligate breaks and is identified by the RHMINBRK option of RHMAP. This approach may be considered analogous to genetic mapping by minimising recombinants (Thompson, 1987). RHMINBRK evaluates and provides several putative marker orders but cannot estimate intermarker distances. These candidate orders can be then compared using maximum likelihood analysis that, for each candidate order, uses the full set of pairwise lod and distance estimates to determine the likeliest marker spacing, and determines the relative likelihood of that order. The RHMAXLIK option of RHMAP (Boehnke *et al.* 1991) performs this analysis, and can use a variety of models to represent the breakage and retention of fragments in the panel. Each of the models assumes that X-ray breakage occurs as a Poisson process along the chromosome, i.e. that there is constant breakage intensity and that there is no interference such that breakage occurs randomly along the chromosome. By making certain mathematical assumptions, a breakage probability, θ, can be converted by the programme to an additive distance , D, using the formula:

$$D = -\ln (1 - \theta) \qquad \text{(Equation 2)}$$

RHMAPPER (Stein, L. 1996, Slonim *et al.* 1997) was developed to handle very large numbers of markers such as are required for the large-scale radiation hybrid mapping of the human genome (for example, Hudson *et al.* [1995]). RHMAPPER uses maximum likelihood methods and an equal retention model to generate maps. However, it is not feasible to evaluate all possible marker orders, due to the computation time required. Therefore, an initial framework is determined by finding well-ordered groups of three markers (triples) and then adding successive markers.

The most important requirements for effective radiation hybrid mapping are accurate marker typing and data entry (Boehnke *et al.* 1991). The large-scale genome mapping efforts have introduced a variety of semi-automated methods developed within their own institutes to reduce these errors and markers are routinely typed in duplicate. Any discrepancies between duplicate typings can then be noted and a further experiment performed if necessary.

The advantages of WG-RH panels are that they permit the mapping of non-polymorphic markers such as genes (type I loci) as well as the polymorphic type II loci. This avoids the need for pedigrees of families with informative meioses and it is relatively inexpensive and efficient to perform. WG-RH mapping has been exploited in a number of mammalian genome mapping projects including mouse, (McCarthy *et al.* 1997), cattle (Womack *et al.* 1997), pig (Yerle *et al.* 1998) rat (Watanbe *et al.* 1999), cat (Murphy *et al.* 1999) and dog (Priat *et al.* 1998).

Radiation hybrid mapping does have some limitations, however. Retention frequencies for markers on the X chromosome will be half the frequency of autosomal makers in WG-RH panels constructed from male donor cells. Therefore, the availability of a WG-RH panel constructed from female donor cells assists mapping chromosome X markers. Chromosome Y markers cannot be mapped on female panels, unless they are present in the pseudoautosomal regions of X. The data from radiation hybrid mapping is derived from the whole genome at a fixed resolution for each panel. It is therefore beneficial to have complementary panels constructed with different amounts of radiation. The human GB4, G3 and TNG panels, for example, were made with 3000, 10,000 and 50,000 rads respectively, and can achieve resolutions of approximately 1Mb, 240kb and 50kb respectively (Stewart *et al.* 1997). In the dog, the recent construction of a second WG-RH panel using 5000 rads (Priat *et al.* 1998) will complement the T72 panel, constructed with 3000 rads. Mapping using these dog panels is at a preliminary stage, so reliable estimates of resolution are not available. However, the 3000 rad

panel can be expected to produce a longer-range, low-resolution map, similar to the GB4 human panel, whereas the 5000 rad panel should give higher resolution.

Disease loci cannot be directly mapped using WG-RH mapping since they require families in which the phenotype can be observed to segregate.

Where type I loci are typed in WG-RH mapping, there may be a problem due to co-amplification from the recipient rodent genome. This is a particular problem with mapping rodent species such as the mouse, where the use of a hamster recipient cell line has resulted in large numbers of primer pairs being rejected for use because of significant co-amplification problems, as might be expected from such closely related species (Jenkins, 1998). Additional limitations of WG-RH mapping include the difficulties in making and culturing the hybrids, and distortions in the maps arising from the effects of certain donor sequences (such as telomeres) on the retention of fragments by the hybrids (McCarthy, 1996).

1.5.2.3 HAPPY mapping

The technique of HAPPY mapping (Dear & Cook, 1993) uses segregation to map markers, but avoids the need to find families as required by linkage mapping. It does not require radiation hybrid panels and thus avoids the problems associated with that technique. The principle is to randomly break genomic DNA by irradiation or mechanical shearing, and dilute the resulting fragments into aliquots containing approximately one haploid genome equivalent. The markers are effectively segregated by the dilution step and are detected using the PCR; linked markers tend to be found together in an aliquot. The map order and distance can be deduced from the frequency with which markers co-segregate, just as in radiation hybrid mapping. The degree of random breakage can be adjusted to generate panels of different resolution as required and relatively small amounts of template DNA are needed to make the panels. Although, like WG-RH mapping, HAPPY mapping cannot be used to map disease loci, it can be used to generate extremely high-resolution maps without gaps. It has been applied to map human chromosome 14, (Dear *et al.* 1998), the whole genome of *Cryptosporidium parvum* (Piper *et al.* 1998) and chromosome 6 of *Dictyostelium discoideum* (Konfortov *et al.* - submitted to Genomics).

1.5.2.4 Physical mapping

Physical mapping is generally taken to mean the mapping of cloned DNA fragments of the genome in a contiguous, overlapping array (or contig). The earliest physical maps were made by comparing the sizes of fragments produced by restriction digestion of clones. Those clones that share many fragment sizes are presumed to overlap, the common fragments being located in the overlapping regions. More recently, physical mapping has made use of the PCR to screen libraries with many different STS markers from the genome of interest. Clones that contain the same STS are presumed to overlap. In this way, not only are the order of cloned fragments established, but also that of the STS markers.

Both these techniques are very labour intensive as a large number of clones must be examined in order to find those which cover the majority of the genome. Even with this high redundancy, the method also will produce maps with gaps because some regions of the genome are resistant to cloning. For example, the recent publication of the sequence of human chromosome 22 used physical mapping to arrange a minimal tiling path of clones that were sequenced; this map, although a major milestone, does contain 11 gaps, amounting to approximately 1Mb of sequence (Dunham *et al.* 1999).

1.5.2.5 FISH mapping

The principal of *In situ* hybridisation is to immobilise chromosomes on a glass slide, DNA fragments, usually cloned DNA, are labelled in some detectable way, and hybridised to the chromosomes, whereupon they bind to the homologous sequence. Detection of the probe allows the location of the fragment of DNA. In early mapping experiments, the probes were labelled with radioisotope and the slides treated with photographic emulsion to visualise the site of hybridisation. This method has largely been superseded by the use of fluorescent dyes (for example, digoxigenin; Zischler *et al.* 1989) and sensitive imaging equipment often in conjunction with sophisticated computer software. Fluorescence *in situ* hybridisation (FISH) has several advantages over probes labelled with radioisotopes. It is a faster process, visualisation being immediate, in contrast to the days or weeks of exposure necessary to detect isotopic probes. The reporter molecules permit greater spatial resolution and are stable for very long periods. The use of fluorescent probes with different emission spectra (i.e. producing different colours) permits more than one probe to be hybridised at a time. Probes for use in FISH are typically derived from cloned fragments, and can be positioned on metaphase chromosomes at a

maximum resolution of between 2 and 3Mb (Trask, 1991). However, higher resolution FISH is possible by hybridising to interphase chromosomes, which are much more extended than metaphase ones, or to artificially extended DNA fibres (fibre-FISH; Florijn *et al.* 1995). The intensity of the signal observed in FISH depends upon the size of the hybridised region. In general, therefore, probes made from small cloned fragments or single PCR products produce weak or undetectable signals. However, repetition of the target sequence within the genome also affects signal strength: a small (1kb) probe hybridising to a unique location in the genome may be hardly visible following FISH, whereas a probe of the same size but derived from a tandemly repeated sequence may produce an intense signal. (If the probe hybridises to a dispersed repeated sequence, then a faint and scattered signal is seen.)

Blood samples provide a reliable method of obtaining white blood cells suitable for stimulation into growth and the subsequent preparation of metaphase spreads for FISH. Cells may be stimulated in whole blood culture or by isolating the lymphocytes for separate culture. The method of choice depends on the species under investigation; in dogs, both methods yield satisfactory results. The cells are stimulated to divide by the addition of mitogens such as pokeweed mitogen (PWM, a crude extract of the roots of *Phytolacca americana*) or phytohaemagglutinin (PHA, a crude extract of the bean seeds of *Phaseolus vulgaris*). Cultures may be treated with agents to increase the number of cells in mitosis at the point of harvesting. This synchronisation can be achieved by a variety of means. If the cells in culture are adherent to a surface, for example fibroblasts, they can be grown to confluence. Then, the effect of crowding causes contact inhibition, which, with the subsequent release by splitting the confluent culture, drives many of the cells into mitosis (Trask, in Birren *et al.* 1999, p306). Alternatively, arrest of cells in metaphase can be achieved by incubation of the dividing cells with a spindle inhibitor, such as colcemid, which prevents the chromosomes from being pulled to the poles during anaphase. Metaphase chromosomes can be prevented from contracting too much by the addition of DNA intercalating agents such ethidium bromide for the final few hours of the culture (Rønne, in Halnan 1989, p15). The cells are then swollen to enable the chromosomes to be studied by incubating them in a hypotonic solution (usually an aqueous potassium chloride solution, 75mM). Repeated rinses in a solution of methanol and acetic acid then fix the chromosomes to preserve their morphology whilst removing the surrounding cytoplasm. Finally, the metaphase chromosomes are fixed to the surface of a glass slide for hybridisation (see section 2.8.2).

Probes may be labelled by two basic methods, direct and indirect. Direct labelling employs the incorporation of fluorescently labelled nucleotides. Directly labelled probes produce sharp signals without the need for secondary detection molecules (and therefore, results are available more quickly); however, the signal may not be as bright as with indirectly labelled probes (Nath & Johnson, 1998). Indirect labelling is the more common technique and involves labelling the probe by the incorporation of modified nucleotide containing a reporter molecule, such as biotin or digoxigenin, by nick translation, PCR or random priming (Korenberg *et al.* 1992). Where nick translation is used, the DNAase activity is optimised to reduce the probe to fragments from 200 to 600bp in size. Such probes are then detected using an affinity molecule that binds to the reporter molecule; for example, biotinylated probes may be detected by avidin conjugated to a fluorochrome such as fluorescein. This signal can be enhanced by subsequent incubation with biotinylated anti-avidin antibody and fluorochrome-conjugated antibody. Prior to hybridisation, the fixed metaphase chromosomes are denatured under conditions that preserve chromosome morphology; the probe is denatured, to allow complementary sequences in the probe and target to anneal during subsequent incubation. It is also necessary to pre-hybridise the metaphase spreads with unlabelled genomic DNA to suppress non-specific hybridisation of the probe, particularly if it contains some repetitive elements dispersed throughout the genome (Trask, 1991). After washing to remove unbound probe (and after treatment with the appropriate affinity molecule, in the case of indirectly labelled probes), signal is observed using fluorescence microscopy. Images can be photographed or captured digitally with a cooled-CCD camera, computer equipment and software.

In addition to visualising the bound probe, it is necessary to visualise the chromosomes themselves so that the location of the probe can be placed in context. Unstained chromosomes may be visualised using phase-contrast light microscopy, which permits chromosome enumeration, analysis of fragile sites and some analysis of heterochromatin content (Halnan, in Halnan 1989). However, methods that are more informative are necessary to reliably identify most chromosomes. Several staining procedures can be used that result in characteristic bands being visualised along the chromosomes (Trask, in Birren *et al.* 1999). The most frequently used method in karyotyping is GTG staining which gives dark (G-positive) bands and light (G-negative) bands (Seabright, 1971) when metaphase chromosomes are treated with trypsin and then exposed to Giemsa stain. The first FISH experiments used Giemsa as a counterstain after the probes had been hybridised and photographed. This method suffered

from a number of drawbacks. In addition to being time-consuming, the extra stages of processing risk damaging the metaphase spread, and the method also requires that the probe signal and banding pattern are imaged independently and then re-aligned afterwards. It is much simpler, more accurate and less time-consuming to stain the chromosomes at the same time as the FISH probes are hybridised. A number of fluorescent stains are available for this purpose, including propidium iodide, and one of the most useful, 4', 6-diamidino-2-phenylindole (DAPI; Cherif *et al.* 1990). DAPI can be applied to the chromosomes as a counterstain following fluorescent probe detection and it produces a banding pattern (Q-banding) almost identical to the G-banding pattern produced by Giemsa. This not only allows specific chromosomes to be identified, but allows the location of FISH probes to be determined relative to the well-characterised G-banding pattern (for example, Solinas-Toldo *et al.* 1995).

Although FISH is commonly used to pinpoint the location of unique sequences in the genome, several variations on FISH rely upon widespread hybridisation of more complex probes. For example, a species-specific probe may be generated from total genomic DNA, total cDNA, or species-specific repetitive elements or by repeat-primed PCR. Such probes will hybridise to multiple sites in the genome of the target species, serving as 'chromosome paints'. For example, the SINE *Alu* is found specifically in primate genomes; primers designed to the ends of the sequence can be used in the PCR to produce primate chromosome paints which can identify the human component from the rodent background of a somatic cell hybrid (Tagle & Collins, 1992).

Paints which identify specific chromosomes from a chosen species can also be prepared by bivariate flow sorting, amplifying and labelling the chosen chromosome (Telenius *et al.* 1992). Such chromosome-specific paints have been produced for chromosomes from a number of species including, man (Cremer, *et al.* 1988, Lichter *et al.* 1988, Pinkel *et al.* 1988), mouse (Rabbitts *et al.* 1995), pig (Langford *et al.* 1992, 1993) and dog (Langford *et al.* 1996, Yang *et al.* 1999). Chromosome paints produced from one species may also be hybridised to chromosome preparations from another species in order to identify the distribution of evolutionarily conserved chromosome segments, a technique known as reciprocal chromosome painting or ZOO-FISH (Scherthan *et al.* 1994). Paints from different chromosomes can be labelled with different fluorophores (or combinations of fluorophores), allowing them to be distinguished in a single hybridisation. This type of probe can be used to identify individual chromosomes and is particularly suited to detecting chromosomal abnormalities that might be

difficult to reliably identify using conventional light microscopy; a panel of paint probes constructed from combinations of fluorophores have been developed that identify each of the 22 autosomes and the sex chromosomes of man, (M-FISH; Speicher *et al.* 1996, spectral karyotyping; Schröck *et al.* 1996) and 19 autosomes and two sex-chromosomes of mouse (Liyanage *et al.* 1996).

Finally, certain repeat elements can be used as FISH probes to identify characteristic regions of chromosomes. For example, α-satellite DNA is characteristic of centromeres, and probes made from this repeat can be used to identify the centromeres in metaphase spreads (for example Modi *et al.* 1988).

1.6 Comparative mapping

Comparison of genome maps is possible where the mapped markers share some detectable homology between the species. Such comparative mapping can reveal the complex pattern of rearrangements that have taken place during the course of evolution. Moreover, comparative mapping between species can allow map information from a better-studied species to be applied to the less well studied. Typically, it is found that gene order is preserved between related species across large segments of chromosomes, a phenomenon known as synteny. Hence, if the order of genes in one species is known to be A-B-C-D-E-F, whilst that of the homologous genes in another species is known to be A'–C'–E'-F', it may be assumed that the segment shows conserved synteny, and the location of B' and D' in the second species can be inferred with some confidence. Such syntenic mapping is particularly valuable in cases where a gene is easier to map in one species than in another. Data from model organisms may not be directly applicable to other, more complex ones, but it may at least suggest likely locations for the gene of interest, or suggest likely candidate loci in the second genome.

Significant genome mapping projects are under way for diverse organisms, some examples of which are included in table 1. These projects generate enormous amounts of data over short periods and consequently traditional means of publishing the data in journals has become somewhat redundant and an inefficient medium. Many of the references are short explanatory articles, which refer to electronic databases accessed via the Internet, and universal resource locators (URL) to the relevant databases have become accepted as citations.

FISH is one of the most powerful tools for comparative mapping, as it provides both long-range positional information and absolute chromosomal assignments for loci or regions. Chromosome-specific paints, when applied to the genome of a different species (reciprocal chromosome painting; Wienberg *et al.* 1990, 1992, Ried *et al.* 1993, Scherthan *et al.* 1996), can quickly identify large chromosomal segments which appear to be conserved, although they cannot guarantee that synteny is preserved within each such segment. Further experiments are therefore required to confirm synteny using specific loci.

Table 1 Examples of genome mapping projects and their Internet addresses.

Caenorhabditis elegans	http://www.sanger.ac.uk/Projects/C_elegans
Cattle, *Bos taurus*	http://www.ri.bbsrc.ac.uk/bovmap/bovmap.html
Chicken, *Gallus gallus*	http://www.ri.bbsrc.ac.uk/chickmap/arkchick
Dog, *Canis familiaris*	http://www-recomgen.univ-rennes1.fr/doggy.html
Haemophilus influenzae	http://www.tigr.org/tdb/mdb/hidb/hidb.html
Human, *Homo sapiens*	http://www.hgmp.mrc.ac.uk
Mouse, *Mus musculus*	http://www.informatics.jax.org
Pig, *Sus scrofa*	http://www.ri.bbsrc.ac.uk/tcagdb.html
Puffer fish, *Fugu rubripes*	http://fugu.hgmp.mrc.ac.uk/fugu/fugu.html
Rat, *Rattus norvegicus*	http://ratmap.gen.gu.se/
Rice, *Oryza sativa*	http://www.staff.or.jp
Saccharomyces cerevisiae	http://genome-www.stanford.edu/Saccharomyces
Sheep, *Ovis aries*	http://dirk.invermay.cri.nz

The first use of reciprocal chromosome painting in the dog was by Werner *et al.* (1997) who used a human chromosome 17 paint to identify the syntenic regions of the dog. They found the human paint was localised to two dog chromosomes that they identified as 9 and 5. To investigate the syntenic relationship more closely, they then used human, mouse and dog probes for loci present on human chromosome 17 to examine the relative arrangement of these markers in the dog using FISH, RFLP analysis and simple sequence repeat polymorphisms. The order of genes on the corresponding dog chromosomes was similar to that in human.

However, of those mapped to chromosome 9, five genes in a block, although in the same order in the dog, were in the opposite orientation (P4HB, GALK, TK1, GH1 & MYL4) and five other genes had a rearranged order in the dog but in the same orientation as in man (BRCA1, RARA, THRAR1, NF1 & CRYBA1). Two other genes located on human chromosome 17 located in the same block, (PMP22 and GLUT4) were found to be located on a separate chromosome (chromosome 5) in the dog.

Another reciprocal chromosome study involving dog and human paints was carried out by Thomas *et al.* (1999). This study refined the work of Werner *et al.* (1997) and demonstrated that human 17p13 corresponds to dog chromosome 5 and human 17p12 - qter corresponds to dog chromosome 9.

Lin *et al.* (1999) used data from research on narcolepsy in dogs to identify and map the gene responsible for canine narcolepsy, hypocretin receptor-2 gene. Hypocretin is a neuropeptide, which prior to the study of Lin *et al.* (1999) was thought to be primarily involved in the regulation of feeding. The work of Lin *et al.* (1999) demonstrated that narcolepsy in dogs was a result of a mutation in the receptor for hypocretin. Hypocretin and it's receptors are consequently being studied as potential candidates for the cause of narcolepsy in humans; this example shows that information from model organisms such as the dog, can be of value in assisting human genetic disease research.

Comparative studies on dog and human by van de Sluis *et al.* (1999) identified a new locus involved in progressive liver disease. Whereas the ATP7B gene mapped to canine chromosome 22q11, CO4107, a microsatellite marker showing close linkage to copper toxicosis mapped to dog chromosome 10q26. A transcribed sequence identified from a CO4107-containing BAC was found to be homologous to a gene expressed from human chromosome 2p16-p13, a region devoid of any positional candidate genes.

Similarly, a mouse Zfy-2 gene was used to screen a bovine cDNA library to obtain clones containing the homologous ZFX gene. The clones obtained were used to locate the bovine ZFX gene to Xq34 and Yp13. The bovine clones were used in both sheep and goat: in sheep it mapped to Xq13 / Yp12 - 13 but in goat, only mapped to Yp12 -13 (Xiao *et al.* 1998).

These examples demonstrate the usefulness of comparative mapping approaches in a variety of mammalian species. The techniques provide information about putative homologies that can be further examined to determine whether a true homology exists between species.

Chromosome X was one of the main subjects of the work described in this study, and is of particular interest in the context of comparative genomics. In order to maintain equal levels of gene expression in male (XY) and female (XX) cells, the majority of the genes on one copy of the X chromosome in female cells are normally repressed, in contrast to autosomal genes which are typically expressed from both copies. Such X-inactivation discourages translocations between X and the autosomes (as the levels of expression of the translocated genes would be altered), a phenomenon known as Ohno's law (Ohno, 1967, 1973). Consequently, the majority of genes present on the X chromosome of one species will be found on that of other species, although not necessarily in the same order.

Comparative cytogenetic studies of the X-chromosomes of a number of mammals have provided insights into the possible evolution of the chromosome from an 'ancestral' X chromosome (Graves *et al.* 1998b). The X-chromosomes of primates, most carnivores, pigs and camels share very similar size, shape, morphology and banding pattern. Those of the Bovidae are much more variable, with rearrangements of the bands and differences in the overall size due to heterochromatic variation or (in an exception to Ohno's law) X-autosomal translocation. Similarly, the X chromosomes of the Cricetidae (including the hamster) and Muridae (mice and rats) have similar euchromatic parts but have differences in the heterochromatin and show X-autosomal translocations (Graphodatsky, 1989 in Halnan 1989).

Molecular studies can provide detailed insights into sex-chromosome evolution. Scherthan *et al.* (1994) used flow-sorted human chromosome paints in order to examine the conservation of synteny in a number of species. They used libraries constructed from human chromosomes 1, 16, 17 and X. These ZOO-FISH experiments showed the X chromosomes of species as diverse as the mouse, fin whale and Chinese muntjac deer contain conserved sequences.

Millwood *et al.* (1997) compared the X-chromosome maps of rat, mouse and humans, and found long stretches of conserved gene order. The genes and markers studied in the Xq regions are all conserved between human and rat, whereas in mouse this region has been disrupted into two. In the Xp region, there are several rearrangements, with three discrete blocks of genes being rearranged in mouse and rat. It appears that the rat X-chromosome shares more conservation with man than does the mouse X.

Comparative mapping studies in other mammalian species and the development of PCR primers from the conserved regions of exons has allowed data from one species to be

directly compared with another (O'Brien *et al.* 1993, Millwood *et al.* 1997, Hu *et al.* 1997). In comparative mapping studies between man and mouse, where 50 of the more than 400 known human X-linked genes have been mapped, homologues of all but three are located on the mouse X-chromosome (Disteche *et al.* 1992). It has been shown that three human X-linked genes, granulocyte macrophage colony stimulating factor receptor (CSF2RA), interleukin-3 receptor, alpha (IL3RA) and voltage gated chlorine channel (CLCN4) are located on different mouse autosomes. CSF2RA is located on mouse chromosome 19 and IL3RA is located on chromosome 14. CLCN4 has been shown to be present on the X chromosome of *Mus spretus*, but located on chromosome 7 of laboratory strains of *Mus musculus*, which indicates a recent (in evolutionary terms) X/autosome translocation in *M. musculus* (Disteche *et al.* 1992, Milatovich *et al.* 1993, Palmer *et al.* 1995, Rugarli *et al.* 1995).

In the pig, three genes (hypoxanthine phosphoribosyltransferase [HPRT], coagulation factor IX [F9] and calcium-binding protein, [CALB3]) were investigated by FISH and RFLP analysis (Hu *et al.* 1997). The three genes (HPRT1, CALB3 and F9) were found to map in the same order on the pig X chromosome as on the human X and their approximate physical locations appeared to be similar.

Although the pseudoautosomal region appears to be quite different from the rest of the X chromosome, conservation of synteny has been shown here. Toder *et al.* (1997) examined three genes CSF2RA, ANT3 and STS located in the human PAR and compared their position in dog and sheep. All three genes were mapped by FISH in both dog and sheep and found to be located in the Xp/Y PAR1 of these species.

In the context of comparative mapping, it is important to remember that some apparent homologies may arise from members of gene-families or pseudogenes. Nevertheless, comparative genomics is a powerful tool, which may help in enhancing the knowledge of dog genes and diseases. A further examination of X-chromosome syntenic relationships between dog and other species is described in section 4.7.

1.7 Dog genome mapping prior to this study

The majority of dog genome mapping prior to this study had been based on genetic linkage mapping. In humans, such maps form a major part of the basis of genetic counselling. In those animals over which man has control of their breeding, this information allows decisions over which animals are allowed to breed.

Linkage mapping requires knowledge of the family structure so that the inheritance of a polymorphic marker can be followed through the pedigree (section 1.5.2). Several groups have used small, limited families for the identification of the mode of inheritance of specific diseases within a population of interest (for example, Deschênes *et al.* 1994). Many well-documented pedigrees are available but dog breeding by its very nature reduces variability by restricting the animals that may be bred from to retain breed characteristics (many bitches may be mated to a few "successful" dogs, i.e. dogs that have won many dog shows). For genome-wide linkage mapping, it is desirable to have more outbred populations in which markers will show more variability. The formation of the DogMap consortium in 1993 expedited the mapping effort by making a reference panel of families available, as shown in figure 2. The DNA from these individuals was distributed among consortium members for marker typing. The DogMap pedigrees consisted of only two generations of animals in which the linkage phase (section 1.5.2.1) could not be determined. Lack of phase information greatly reduces the power of linkage mapping. Nevertheless, the result of this work was the generation of the first genetic linkage map of the dog genome (Lingaas *et al.* 1997). The 94 loci typed consisted of 85 dinucleotide repeats, three tetranucleotide repeats, one microsatellite located in the canine von Willebrand gene (Shibuya *et al.* 1994) and five protein polymorphisms. The protein polymorphisms consisted of four blood plasma proteins, transferrin (TF), alpha-1-protease inhibitor (PI), alpha-1B-glycoprotein (A1BG) and apolipoprotein A4 (APOA4), and an erythrocyte enzyme, superoxide dismutase 1 (SOD1). Two-point linkage analysis of these markers produced 16 linkage groups from 43 of the markers, the remainder being unlinked. This relatively small number of linkage groups and high number of unlinked markers were a consequence of only having 94 markers available to map the entire dog genome and possibly, also the limitations of the reference family used to generate the map.

Late in 1998, Cornell University released to the general mapping community a selection of three generation, phase-known families produced from a number of different breeds of dog.

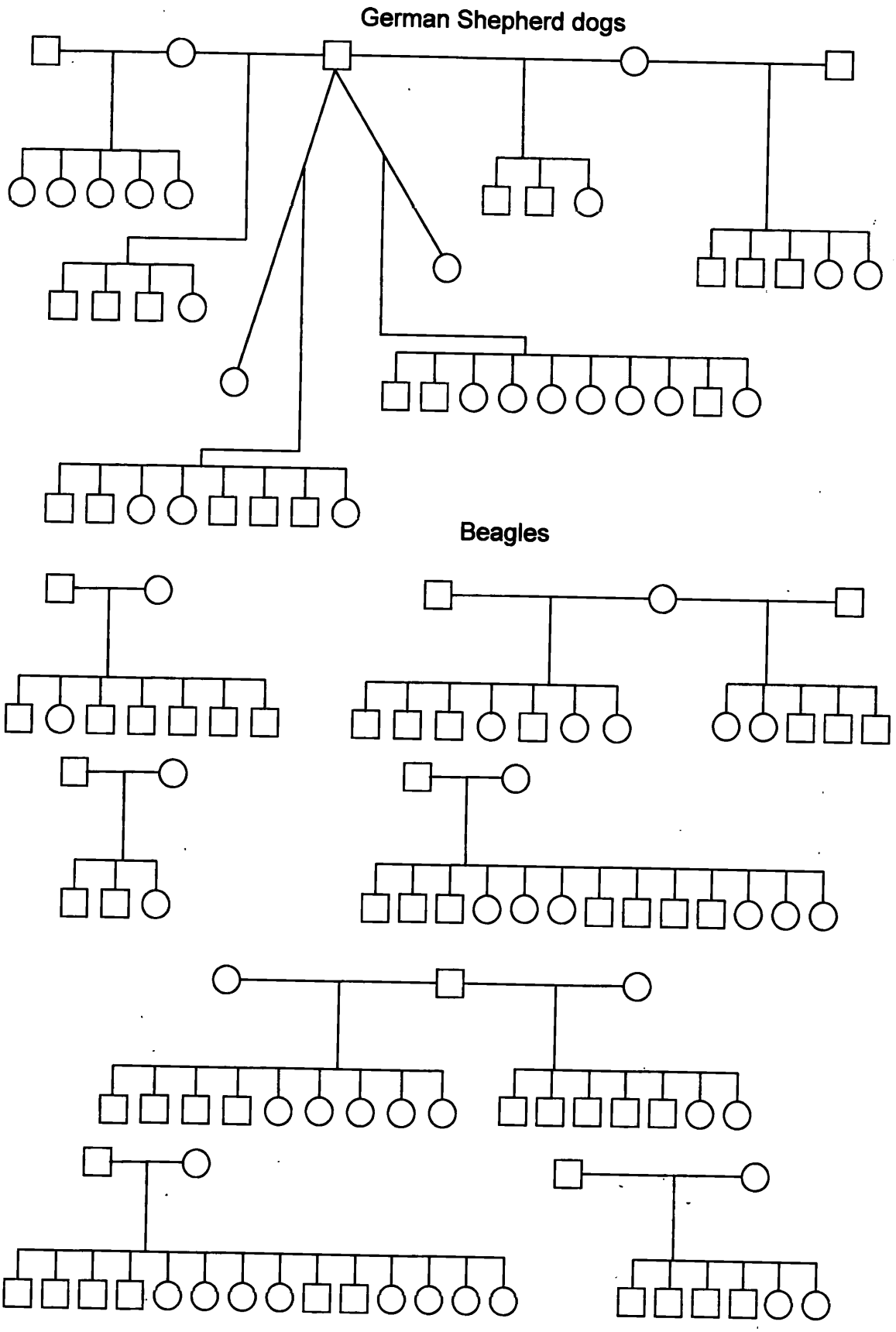


Figure 2 Pedigree structure of the DogMap families.
The DogMap families consist of two breeds, German shepherd dogs in one family, and small families of Beagles. In total, there are 130 individuals, 42 German shepherd dogs and 88 Beagles.

This overcame many of the limitations of the DogMap families and improved the mapping effort. Only a sub-set of these animals was made publicly available, as shown in figure 3. This set of animals is composed of the following breeds: (CF = Cornell family)

CF4 = Mixed bred Poodle – Beagle x Mixed bred Poodle-Norwegian Elkhound,

CF5 = Mixed bred Poodle – Beagle x Mixed bred Poodle-Norwegian Elkhound

CF6 = Mixed bred Poodle – Beagle x Purebred Miniature Poodle x Toy Poodle

CF7 = Mixed bred Poodle – Beagle x Purebred Beagle

It is apparent that this still does not represent the ideal reference family since many of the animals are actually from the same breed, limiting the variability seen in polymorphic markers. However, it is three-generation, phase known and sufficient quantities of DNA form each animal have been obtained to enable many groups to type markers on the common resource.

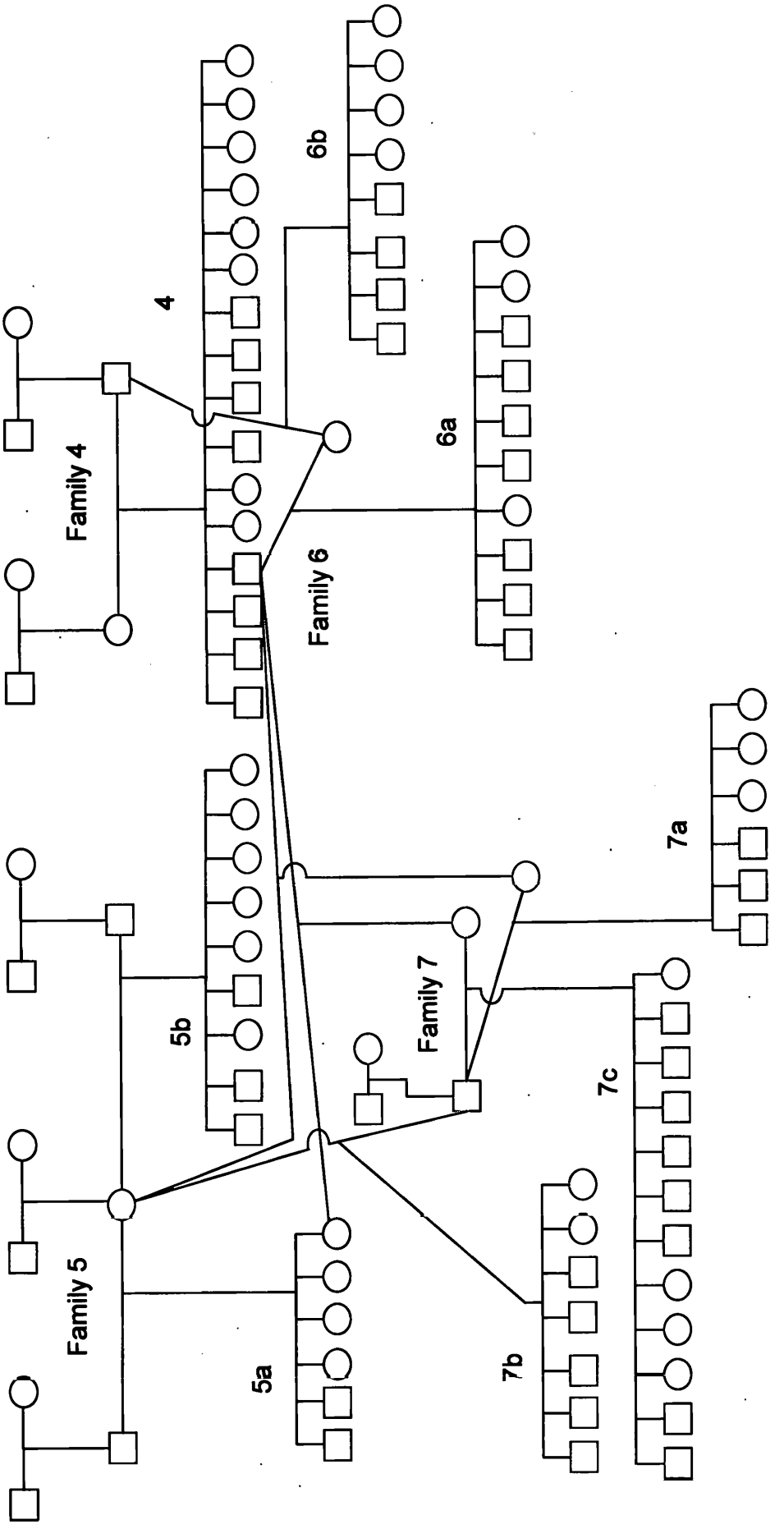
Two linkage maps have been published using members of the Cornell families (Mellersh *et al.* 1997, and Neff *et al.* 1999). The map of Neff *et al.* (1999) is a second-generation map, extending that of Mellersh *et al.* (1997). The most recent of these maps used 218 animals from the Cornell families, in addition to a further 29 related animals. (This represents a larger set of animals than that available to the AHT.) The markers used in these maps included 15 type I loci and 261 microsatellites to generate 39 autosomal and one X-linked linkage group. Unfortunately, the lack of chromosomally assigned markers has been a recurring problem in dog genome mapping, and made it difficult to assign the autosomal linkage groups to specific chromosomes. Recently, progress has been made to address this. Werner *et al.* (1999) used type I and type II markers to integrate with the previous dog maps. They have assigned 25 linkage groups to 14 dog chromosomes by a combination of FISH and genetic linkage mapping.

The chromosome X linkage group of Neff *et al.* (1999) was composed of two linked markers, androgen receptor (AR) and phosphoglycerate kinase 1 (PGK1) with an estimated 23cM between them. Three other markers could not be placed relative to AR and PGK1 but are known to be located on the X chromosome from their sex-linked patterns of segregation: choroideraemia (CHM), clotting factor VIII (F8c) and a microsatellite, C0X.314 (Ostrander *et al.* 1993).

Figure 3 Pedigree structure of the Cornell reference families available to the AHT.

Family 4 (CF4) is shown in red, family 5 (CF5) is shown in green, family 6 (CF6) is shown in blue and family 7 (CF7) is shown in yellow. In total DNA from 94 of the 95 members of the pedigree were available to the AHT.

CF4 comprises one set of parents plus their respective grandparents and 16 offspring (eight males and eight females)
CF5 comprises one mother producing two litters with two different fathers, with the respective grandparents for each. CF5a consists of six offspring, two males and four females. CF5b consists of nine offspring, three males and six females.
CF6 uses the father of CF4 and one of the CF4 male offspring to mate with one female and produce two litters. The grandparents for CF6 are from CF4. CF6a consists of ten offspring, seven males and three females. CF6b has eight offspring, four males and four female.
CF7 has two fathers and three mothers producing three separate litters. The grandparents for CF7 are from families CF4 and CF5. CF7a has six offspring, three males and three females. CF7b uses the mother of CF5 to produce seven offspring, five males and two females. CF7c consists of 12 offspring, eight male and four female.


Family 4
4
6b
6a
Family 6
Family 5
5b
5a
Family 7
7a
7c
7b

None of the pre-existing dog linkage maps (Lingaas *et al.* 1997, Mellersh *et al.* 1997, Neff *et al.* 1999) had a linkage group assigned to chromosome 1. Producing markers known to be located on this chromosome, which could be mapped and integrated into the existing maps, would thus define which of the earlier linkage group or groups were located on chromosome 1.

A small number of individual markers had been mapped to dog chromosomes using FISH by several groups (Deschênes *et al.* 1994, Dutra *et al.* 1996, Fischer *et al.* 1996, Guevara-Fujita *et al.* 1996, Werner *et al.* 1997, Dolf *et al.* 1997a, 1997b, 1998, Schelling *et al.* 1998a, 1998b, Schläpfer *et al.* 1998). However, the difficulties in correctly assigning the chromosomes means that these data may be inaccurate, particularly as they were all carried out before the dog karyotype was standardised (Breen *et al.* 1999b).

The generation of the first dog whole-genome radiation hybrid panel (T72) by Dr. L McCarthy made publicly available by Research Genetics Inc. (Huntsville, AL) and the WG-RH_{5000} panel of Priat *et al.* (1998) have permitted the mapping of markers without the need for reference pedigrees of dogs. Priat *et al.* (1998) utilised 218 type I loci and 182 type II microsatellite markers to generate a framework WG-RH map of dog. This map was comprised of 57 linkage groups (although 13 of these were composed of only 2 or 3 markers). Eight groups were assigned to individual autosomes by the inclusion of markers previously assigned to a chromosome, and one group to chromosome X by the inclusion of markers already mapped to this chromosome (PGK1 & CHM, by Deschênes *et al.* 1994). As this panel was created using a male dog, it was possible for SRY to be typed on the panel, although no other markers were found to be linked to it and hence located on chromosome Y. It should be noted that Priat *et al.* (1998) only typed markers once and not in duplicate, as has been shown to produce accurate maps in the human genome (P. Deloukas, personal communication). The chromosomal assignments presented in this paper, as mentioned above (section 1.5.2.5), may be inaccurate due to the difficulty of identifying the dog autosomes. The application of chromosome paints in conjunction with markers from each linkage group would provide a definitive chromosomal location. More recently, Mellersh *et al.* (2000) have utilised an integrated approach to produce an enlarged and combined genetic linkage and WG-RH map of the dog genome. These authors also did not assign markers directly by FISH but used pre-existing FISH mapped markers in their linkage map and, in fact, found inconsistencies with previously published markers and their own results (for example, IGH had previously been mapped to chromosome 4 [Dutra *et al.* 1996] whereas Mellersh *et al.* [2000] found it linked to a marker assigned to chromosome 8 by

Werner *et al.* [1999]). The map of Mellersh *et al.* (2000) typed 217 markers on both the complete Cornell reference families and WG-RH_{5000} panel, thus providing duplications of typings (although they did not use duplicate typings on the WG-RH_{5000} panel itself). This approach gave reliable results; they found only one discrepancy between the markers mapped by both methods.

1.8 Present study

The object of the work presented in this thesis was to increase knowledge of the canine genome and, in particular, of the two largest chromosomes, X and 1. The pre-existing linkage maps had no linkage groups assigned to dog chromosome 1 and only five markers had been mapped to chromosome X. Producing maps of markers with known physical locations to integrate with the existing maps would be extremely valuable to the dog genome mapping effort. The aim was to use a combination of approaches rather than a single approach, both to increase the accuracy and robustness of the resulting maps and to provide the best opportunities for integration with other data.

Microsatellite markers derived from chromosome-specific libraries were chosen as the basis for this approach. They can be mapped both by genetic linkage and radiation hybrid methods, and the clones from which they originate can be mapped using FISH. The features of microsatellites such as their relatively high frequency, Mendelian inheritance and polymorphism have made them extremely valuable in the generation of linkage maps of many species including man (Dib *et al.* 1996), mouse (Dietrich *et al.* 1996), rat (Jacob *et al.* 1995), cattle (Barendse *et al.* 1994, 1997, Bishop *et al.* 1994), pig (Rohrer *et al.* 1994), sheep (Crawford *et al.* 1995, Galloway, *et al.* 1996), horse (Breen *et al.* 1997), goat (Vaiman *et al.* 1996) cat (Menotti-Raymond *et al.* 1999) and dog (Lingaas *et al.* 1997, Mellersh *et al.* 1997, Neff *et al.* 1999). Microsatellite markers are recognised as being eminently suitable for mapping in the dog (Ostrander *et al.* 1993). They have been shown to be useful within the relatively inbred dog populations by studies on polymorphism in a variety of breeds (Holmes *et al.* 1993a, 1993b, 1995, Rothuizen *et al.* 1994, Fredholm & Winterø, 1995, Zajc *et al.* 1997).

Mapping the dog chromosome X was a priority, in part because whole genome screens carried out prior to this study had failed to obtain and map any microsatellites markers on

chromosome X. In the mouse (Dietrich *et al.* 1996), man (Dib *et al.* 1996) and the rat (Jacob, *et al.* 1995) paucity in the number of microsatellites located on chromosome X has been reported. In addition, the average heterozygosity of microsatellite markers on X in man (Dib *et al.* 1996) and mouse (Dietrich *et al.* 1996) was lower than the other chromosomes. It was therefore expected that the dog X markers would similarly show a reduced degree of polymorphism.

A further reason for interest in chromosome X is the high degree of synteny observed between the X-chromosomes of other species. An X-chromosome map of dog was therefore expected to be of particular relevance in comparative mapping, which may help elucidate chromosome evolution. Moreover, chromosome X is relatively disease rich: the dog is afflicted by several X-linked diseases, some of which (for example, X-linked severe combined immunodeficiency disease, haemophilia A & B and muscular dystrophy) are models of human disease. Finally, chromosomes X and 1 are the largest dog chromosomes, making them relatively easy to identify cytogenetically, and the practicability of generating good quality pools of these chromosomes by flow sorting had already been demonstrated (Langford *et al.* 1996).

Integration of the data from the different dog maps available now and in the future is very important. In addition, the ability to compare maps between species is of great value for evolutionary studies and positional candidate gene approaches. To this end, markers for type I loci and from existing maps were investigated in this study, to supplement the microsatellite markers. The development of primers from the conserved regions of exons of mammalian species has allowed data from one species to be directly compared with another (O'Brien *et al.* 1993, Millwood *et al.* 1997, Hu *et al.* 1997). Several X-linked genes with these conserved primers have been made available and used in this study. They included comparative anchor tagged sequences (CATS, Lyons *et al.* 1997); universal mammalian sequence-tagged sites (UM-STS) (Venta *et al.* 1996) and traced orthologous amplified sequence tags (TOAST) (Jiang *et al.* 1998). These primers were examined for suitability for linkage mapping and radiation hybrid mapping. A canine cosmid library in pWE15 is commercially available (Stratagene) and has been used in the dog genome project at the Animal Health Trust to identify a cosmid marker on every dog chromosome (Breen *et al.* 1999a). These clones have been screened for the presence of poly (CA) microsatellites (Dickens *et al.* 1999) and have been chromosomally assigned by FISH (Breen *et al.* 1999a). Two of these clones which mapped to chromosomes X and 1 had previously been sub-cloned and sequenced (N. Holmes, personal communication); a further six were end-sequenced in this study. Markers derived from these clones were included

in this study.

2 Materials and Methods

2.1 Flow sorting of canine chromosomes

Dog mononuclear cells were prepared as follows under sterile conditions, to produce cells arrested in metaphase. A healthy female Irish setter dog was bled from the jugular vein and 80 - 100ml of blood collected into 50ml centrifuge tubes (Sterilin) in the presence of 15U/ml preservative-free heparin (Monoparin, C. P. Pharmaceuticals). Separation gradients were prepared consisting of 2.5ml Histopaque 1119 (Sigma) overlaid with 2.5ml Histopaque 1077 (Sigma) in a 15ml centrifuge tube (Falcon). Four and a half millilitres of blood were overlaid onto each of 20 separation gradients and then spun for 30 mins at 600g (Centaur, MSE). Lymphocytes were collected from the serum / Histopaque 1077 interphase, pooled, gently resuspended in 5ml RPMI-1640 and spun for 10 minutes at 200g. Cells were resuspended in 1 - 2mls of RPMI-1640 and the cell count determined by examining 6µl of cell suspension on a Neubauer haemocytometer under x40 magnification. Approximately 1x 10^7 cells were resuspended in 40ml of RPMI-1640 containing 20% v/v foetal calf serum (FCS, Gibco BRL), 1% v/v penicillin / streptomycin (5000 iu/ml each; Gibco BRL) and 1% L-glutamine (2mM final conc. Gibco BRL). Phytohaemagglutinin (HA15, Murex Diagnostics) and pokeweed mitogen (Gibco BRL), were added to a final concentration of 1% v/v each, and the cultures were incubated at 37°C, 5% CO_2 for 72 hours. Colcemid (100ng/ml, Gibco BRL) was added and the incubation continued for 16 hours to arrest the cells in metaphase. The culture was decanted into a 50ml centrifuge tube and spun for 10 minutes at 200g at room temperature. The supernatant was removed and the tube drained for 3 - 5 minutes. The pellet was resuspended in 10ml 75mM KCl containing 0.2mM spermine tetrahydrochloride (Sigma) and 0.5mM spermidine trihydrochloride (Sigma) and incubated at room temperature for 15 minutes to swell the cells. The suspension was centrifuged for ten minutes at 400g, the supernatant was removed and the pellet resuspended in 2ml of modified polyamine buffer (Siller & Young 1981, Carter *et al.* 1994, appendix 1). The preparation was placed on ice for 30 minutes then vortexed for 15 seconds. An aliquot of 6µl of the suspension was stained with 1µl ethidium bromide (10mg/ml) and immediately examined under ultraviolet fluorescence to confirm that the chromosomes had been released from the cells. After the addition of 20µl sodium azide (15% w/v), the preparation

was spun for two minutes at 100g to pellet cell debris and nuclei, leaving the chromosomes in suspension. The supernatant was collected and remaining debris removed using a 70μm nylon filter cell strainer (Falcon). Filtered chromosome suspension was stained at 4°C for 2.5 hours with 150μl of chromomycin A3 (0.15mg/ml final conc. Sigma), 15μl of Hoechst 33258 (7.5ng/ml final conc. Sigma), supplemented with 7.4μl 1M $MgSO_4$. Sodium citrate (10mM final conc.) and sodium sulphite (25mM final conc.) were added ten minutes prior to sorting to stabilise the chromosomes. The dog flow karyotype was analysed using a dual-laser flow sorter (Elite ESP, Coulter Electronics) and peaks corresponding to dog chromosomes 1 and X were identified by reference to previous dog chromosome flow karyotypes (Langford *et al.* 1996). The computer software was used to isolate the peaks and approximately 200,000 events for each chromosome were collected into two 0.2ml tubes (Perkin Elmer).

2.2 Small Insert libraries - preparation

Flow-sorted chromosomes (approx. 200,000 events for either chromosome X or 1) were supplemented with 50 μl 1% sodium *N*-lauroylsarcosine (Sigma) and proteinase K (32μg/ml final concentration; Sigma) then digested at 42°C overnight. Proteinase K was inactivated by adding phenylmethylsulphonyl fluoride (40μg/ml, Sigma) and incubating at room temperature for 40 minutes. Chromosomal DNA was precipitated by adding one-tenth volume of 3M sodium acetate, pH5.2, and two volumes of absolute ethanol prior to incubating overnight at -20°C. DNA was pelleted using a mini-centrifuge (Eppendorf) at room temperature for 15 minutes at approximately 10,000g. The supernatant was removed and the pellet de-salted by washing with 70% ethanol, respun (seven minutes), supernatant removed and the pellet air dried for five minutes before dissolving in 17μl of TE buffer, (10mM Tris-HCl pH8, 1mM Na_2EDTA), to give approximately 3ng/μl. (Langford *et al.* [1996] estimated the dog chromosomes X and 1 to contain 137 and 135Mb of DNA respectively. One megabase of DNA has a mass of 10^{-15}g; hence, 200,000 chromosomes contain 50ng of DNA.) The tubes were kept at room temperature for two hours and agitated occasionally to resuspend the DNA.

Chromosomal DNA was digested by adding 2μl reaction buffer 2, 10 x concentrate (NEB) and 0.5μl *Hin*dIII (20U/μl, NEB) then incubating at 37°C for two hours. Following

digestion, the enzyme was inactivated at 65°C for 30 minutes and the digests subsequently kept on ice.

The selected vector, pBluescript II SK (+) (Stratagene) was digested with *Hin*dIII (20ng vector, 1U *Hin*dIII [NEB], reaction buffer 2 [NEB], 37°C for two hours) and treated with calf intestinal alkaline phosphatase (Sigma) according to the manufacturer's instructions. This treatment dephosphorylated the 5' end of the vector thus preventing self-ligation. Fifteen nanograms (5μl) of digested chromosome X- or 1- enriched DNA were mixed with 5ng (1μl) of the digested vector. To this mixture, 1μl 10x reaction buffer (NEB), 0.5μl T4 ligase (400U/μl; NEB) and 6.0μl TE buffer were added, the ligation mixture incubated at 16°C overnight and subsequently placed on ice.

A 40μl aliquot of electrocompetent *E. coli* XL-1 Blue (Stratagene) cells was thawed on ice. One microlitre of the ligated DNA preparation was added to the cells and electroporated at 200Ω, 25μF and 1.8kV (Gene Pulser, Biorad) using 0.1cm cuvettes (Biorad). One millilitre of LB broth (appendix 1) was added and the sample transferred to a 50ml tube, then incubated at 37°C for 45 minutes. Fifty microlitres of culture were spread onto an LB agar plate (appendix 1) containing 50μg/ml ampicillin, 40μl 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal; 20mg/ml; Stratagene) and 4μl isopropylthiogalactoside (IPTG; 200mg/ml; Stratagene) then incubated overnight at 37°C. The remaining cells were supplemented with glycerol to a final concentration of 50% and stored at -70°C. A further four rounds of electroporation were carried out to give a total of 5ml transformed cells. Recombinant colonies were picked by hand using sterile cocktail sticks into 96 well microtitre plates (Corning) containing 150μl LB broth with 50μg/ml ampicillin (LB amp) and 7.5% glycerol then incubated again at 37°C overnight. In addition, eight 22x22cm (Bioassay plates, Nunc) LB amp agar plates containing X-gal (20mg/ml) and IPTG (200mg/ml) were inoculated with 0.5ml of transformed cells for each chromosome library. Following overnight incubation at 37°C, recombinant colonies (blue) were picked into 384 well microtitre plates using a BioPick robotic colony picker (BioRobotics) or by hand into 96 well plates. The microtitre plates contained 70μl or 150μl LB amp and 7.5% glycerol respectively and were subsequently stored at -70°C.

2.3 Small insert libraries - screening for microsatellites and preparation of clone DNA

Hybond N+ nylon membranes (22x22cm, Amersham) were placed onto large LB agar plates containing 50μg/ml ampicillin. Inocula from the microtitre plates were transferred to the membranes using a steel-pinned replicator and the plates incubated overnight at 37°C. The membranes were carefully removed from each agar plate, placed onto 3MM paper (Whatman) soaked in denaturing solution (appendix 1) and left for five minutes. They were subsequently transferred to 3MM paper soaked in neutralising solution (appendix 1), left for five minutes and then rinsed in a solution of 0.5x SSC, 0.1% SDS to remove cellular debris prior to drying between sheets of 3MM paper.

Polynucleotide $(dA\text{-}dC)_n.(dG\text{-}dT)_n$ probe (Pharmacia) was labelled with α-^{32}P-dATP (Amersham) using the Klenow fragment of DNA polymerase 1 (Gibco BRL) as follows

50ng polynucleotide probe

10μl 5x Klenow buffer (Stratagene)

5μl Klenow fragment of DNA polymerase 1 (50 Units)

5μl α-^{32}P-dATP (10mCi/ml)

made up to 50μl with MQ water

Reactions were incubated for two hours at 14.5°C and terminated by the addition of 10μl 0.5M EDTA. Columns containing Sephadex G-50 (Pharmacia) were washed with 1x SSC and the reaction mix eluted through the column with 1x SSC (appendix 1) to remove unincorporated label according to the manufacturer's instructions. The eluant was collected into fractions of three drops and each fraction examined with a Geiger counter to identify those containing the labelled probe; these fractions were pooled. The pooled eluant was denatured for three minutes at 100°C and placed on ice. The nylon membranes were prehybridised for approximately 30 minutes at 65°C in 20ml modified Church's buffer (appendix 1; Church & Gilbert, 1984) in a rolling bottle (Hybaid). The denatured probe was added to the prehybridised membranes, which were incubated at 65°C overnight (16 - 18 hours). Membranes were initially washed at low stringency (42°C, in three changes of 0.1% SDS, 0.5x SSC over one hour), then mounted on 3MM paper. They were covered with cling-film and exposed to X-ray film (RX 100nif, Fuji) at -70°C for 18 hours with an intensifying screen. This gave an image with a high background to facilitate the correct orientation of the positive hybridisation signals with the

corresponding wells of the microtitre plates. Membranes were rewashed as before but at 65°C to remove the background signal, then re-exposed to identify the positive clones.

The positions of putative microsatellite containing clones were marked and 10μl removed from each of the corresponding glycerol stocks to inoculate 1ml LB amp broth. After overnight incubation at 37°C, plasmid DNA was isolated by alkaline lysis using a commercial kit (Plasmid Mini Kit, Qiagen) following the manufacturer's instructions without modification. The resulting DNA pellets were resuspended in 30μl of TE buffer. Fresh glycerol stocks of each positive clone were prepared by pelleting a 1ml overnight culture of each clone in a mini-centrifuge (Eppendorf) at 2800g for three minutes and resuspending in 0.5ml LB amp, 0.5ml glycerol. These were stored at –70°C and used to prepare further DNA as required.

The amount of DNA obtained from each preparation was determined by comparison with lambda DNA quantitation standards (Gibco BRL) after electrophoresis on 1% agarose gels. The gels were stained with ethidium bromide (0.5ng/μl, Gibco BRL) and examined under ultraviolet light using a transilluminator (UVP). Images were captured using a video camera (Mitsubishi) and photographs taken using a video copy processor (Mitsubishi).

2.4 Small insert libraries - determination of insert size

The average insert sizes of the putative microsatellite-containing clones from the small insert libraries were determined by digestion of a sample of the clones with *Hin*dIII in a 10μl reaction, (approx. 50ng DNA, 1x reaction buffer 2 [NEB], 5U *Hin*dIII [NEB] 37°C, two hours). Digested DNA was examined by gel electrophoresis (8 V/cm) on a 1% agarose gel (NuSieve, Flowgen) in 1x TAE, alongside 1kb size-standard ladder (Life Technologies). The gels were stained with ethidium bromide (0.5ng/μl, Gibco BRL) and examined under UV light using a transilluminator (UVP). Images were captured using a video camera (Mitsubishi) and photographs taken using a video copy processor (Mitsubishi).

2.5 Small insert clone sequencing

The general strategy for sequencing the ends of the small insert clones and, where necessary, for obtaining additional internal sequence is outlined in figure 4.

The sequencing method used employs the chain termination technique of Sanger *et al.* (1977) and fluorescent dye terminator chemistry (Smith *et al.* 1986). Sequencing reactions were performed using the ABI Prism dye terminator sequencing kit with AmpliTaq DNA polymerase, FS, (both from Perkin Elmer), following the manufacturer's protocols except that reaction volumes and content were reduced by half (from 50µl to 25µl). Approximately 200ng of purified plasmid DNA were used in each 25µl reaction with 1.6pmol of the appropriate sequencing primer. Reactions were carried out in a thermal cycler (Tetrad, MJ Technologies) under the following conditions:

96°C	30 seconds}	
50°C	15 seconds}	x24 cycles
60°C	4 minutes }	
4°C	5 minutes	

Denaturing polyacrylamide (4%) sequencing gels were prepared according to the manufacturer's instructions (Perkin Elmer) with the recommended reagents (Amresco). Prior to loading, 4µl sequencing loading buffer (appendix 1) were added to each sample and mixed thoroughly. The samples were denatured at 95°C for two minutes and placed on ice. Two microlitres of sequencing product were loaded onto the gel. Gel electrophoresis was carried out in 1x TBE (appendix 1) as recommended, for seven hours. Analysis of the sequencing products was performed using the manufacturer's software (Sequence analysis 2.1.1, ABI100, Perkin Elmer).

Initial sequencing was performed using the T3 and T7 primers (appendix 1). In those cases where these reads did not extend into the microsatellite, and where microsatellite-flanking sequence was required for subsequent linkage mapping, additional sequencing was performed using each of the microsatellite primers DRP1 - 6 in turn (appendix1). (The primers each consists of a 10-mer dinucleotide repeat followed by a unique base; one of the six is expected to prime successfully on any microsatellite that is a perfect repeat where $n \geq 10$). This microsatellite-primed sequence was then used to design a unique reverse primer from which to sequence back through the repeat and into the opposite flanking sequence, as illustrated in figure 4.

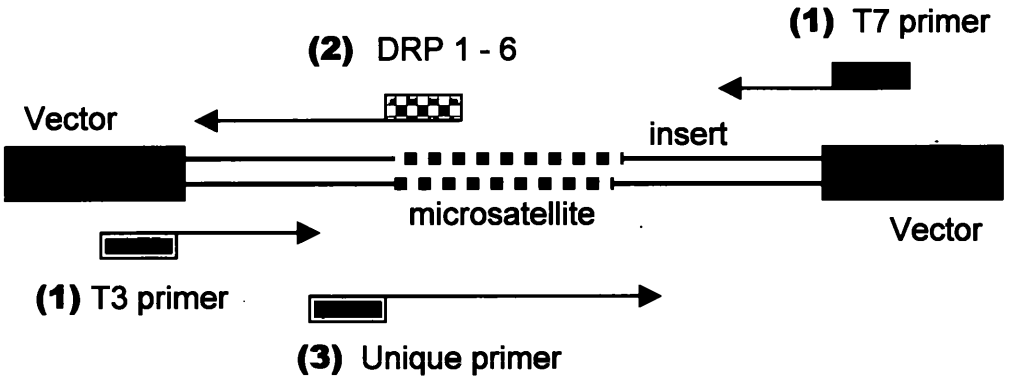


Figure 4 Sequencing strategy for microsatellite-containing pBluescript clones.

Initial sequencing was performed using the T3 and T7 primers complementary to the vector ends (1). Where these sequences failed to extend through the microsatellite (and where microsatellite flanking sequence was needed), further sequencing was tried using the repeat primers DRP1-6, one of which is expected to yield sequence from any given microsatellite (2). In these cases, a unique primer based on the DRP-primed sequence was used to read in the reverse direction to obtain the opposite flanking sequence (3).

2.6 Cosmid clones - screening for microsatellites and preparation of clone DNA

Clones from a genomic canine cosmid library (Stratagene) were replica plated onto LB amp agar (appendix 1) essentially as described above, except that they were grown directly on the agar and then colony lifts were carried out to transfer the DNA onto nylon filters (Hybond N+, Amersham). The cosmid DNA on the filters was denatured and neutralised as described in section 2.3. Identification of clones containing microsatellites was carried out as described previously. The positive clones were isolated and DNA prepared using standard alkaline lysis methods with a commercially available kit according to the manufacturer's instructions (Plasmid Mini Kit, Qiagen). Cosmid clones have a low copy number and initial yields of DNA were low (<100ng in total). DNA was obtained at higher yield by splitting the 15ml culture volume into two tubes and preparing the DNA separately. At the resuspension stage, each pellet was resuspended in half the final volume and the two samples pooled. The DNA was quantitated as described above. The average size of the inserts in this library was not determined in this study but according to the supplier (Stratagene) is in the order of 40kb; it is therefore possible that each clone could contain more than one microsatellite repeat.

2.7 Cosmid clones - end sequencing

Cosmid clone end sequences were derived using the BigDye sequencing kit (Perkin Elmer) according to the manufacturer's instructions without modification. This method can compensate for the reduced amount of template produced by low-copy number vectors such as cosmids, as it uses two dyes instead of one. The nucleotide is linked to a fluorescein donor dye, e.g., 6-carboxyfluorescein (6-FAM) dye, which is then conjugated to a dichlororhodamine (dRhodamine) dye, which fluoresces. The excitation maximum of each dye label is that of the fluorescein donor, and the emission spectrum is that of the dRhodamine acceptor. The donor dye is optimised to absorb the excitation energy of the argon ion laser in the sequencer, which therefore affords very efficient energy transfer between the donor and acceptor dyes. The increased energy output of the two dyes together means that a smaller quantity of template is required than in the standard sequencing protocol. Cosmid sequencing with this method

requires 500ng template DNA per sequencing reaction. The T3 and T7 vector primers were used at 3.2pmol in a 20μl total reaction volume. The sequencing gels were run and analyses performed as for the small insert library clones (section 2.5).

2.8 Fluorescence *in situ* hybridisation (FISH)

2.8.1 Probe preparation

Clones were labelled with the haptens digoxigenin or biotin using nick translation. A labelling reaction mixture was prepared, on ice as follows

1μg clone DNA

5μl 10x FISH nick translation buffer (appendix 1)

5μl of 10x dNTP (appendix 1) mix (containing either biotin-16-dUTP [0.1mM; Boehringer] or digoxigenin-11-dUTP [0.1mM; Boehringer])

2μl deoxyribonuclease 1 (0.0008U/μl final concentration; Gibco BRL)

1μl DNA polymerase 1 (0.2U/μl final concentration, Gibco BRL)

Made up to a final volume of 50μl with water

This was incubated at 14.5°C for one hour; 5μl were removed from the labelling reaction and analysed by electrophoresis of the samples on a 2% agarose gel to ensure that fragments of around 500bp had been produced. If not, then further incubation and analysis were carried out. When appropriately sized fragments had been produced, the reaction was terminated by the addition of 5μl 0.5M EDTA.

The labelled DNA was precipitated by the addition of one-tenth volume of 3.5M ammonium acetate and three volumes of ice-cold absolute ethanol. The mixture was incubated at -20°C overnight and the DNA pelleted by centrifugation at 10,000g (Eppendorf) at 4°C for 30 minutes. The pellet was washed in 500μl ice cold 80% ethanol, respun as before and the DNA air-dried. Labelled probes were resuspended in 6 μl MQ water to give a final concentration of 150ng/μl. A microlitre was removed from each sample and run on a 2% agarose gel with quantitation standards derived from known concentrations of sonicated salmon sperm DNA (Gibco BRL) to check the concentration of precipitated product. The labelled DNA samples were stored at -20°C until required.

2.8.2 Chromosome preparation

Canine peripheral blood was collected and cultures of approx. 10^7 isolated lymphocytes were prepared and cultured in 40ml RPMI-1640 supplemented with 20% foetal calf serum, 1% penicillin/streptomycin (5000 iu/ml each) and 1% L-glutamine (2mM final conc.). The cultures were incubated for 70 hours at 37°C in 5% CO_2. Twenty microlitres of bromodeoxyuridine (BrDU, 20mg/ml, Sigma) were then added to each flask followed by three hours incubation. Then 40µg of ethidium bromide (Gibco BRL) were added to each flask followed by one hour incubation. Finally, 200µl of colcemid (50ng/ml, Gibco BRL) were added followed by incubation for a further hour. Each culture was decanted into a 50ml centrifuge tubes (Falcon) and pelleted at 200g for ten minutes at room temperature. The supernatant was removed and the pellets gently resuspended in the residual medium. Potassium chloride (75mM) at 37°C was added dropwise to a total volume of 6 - 8ml while gently agitating the cells. After ten minutes at room temperature, the tubes were spun at 200g for ten minutes, the supernatant drawn off leaving approximately 1ml of liquid and the pellets gently resuspended. Ice cold, freshly prepared fixative (appendix 1) was added dropwise to each tube until the pellets were resuspended in approximately 8ml. The tubes were spun at 80g for ten minutes, the supernatant removed and the pellets resuspended in another 8ml of fixative. This process was repeated once more and the fixed metaphase preparations stored at -20°C until required.

Prior to preparing metaphase spreads, glass slides (BDH) were cleaned in 2M HCl, washed in MQ water and dried with lint-free tissue. Fixed metaphase preparations were spun at 160g for ten minutes, the supernatant discarded and the pellet gently resuspended in the residual liquid (typically <0.5ml). Seven microlitres of the metaphase suspension were removed and dropped onto a cleaned slide with a pipette (Gilson Pipetman); breathing on the slides immediately before dropping the suspension assisted chromosome spreading. The prepared slides were stored in dust free conditions for at least three days at room temperature before use.

2.8.3 Hybridisation

For each experiment DNA from two clones, one labelled with biotin and the other with digoxigenin were prepared for hybridisation by mixing the following

1µl (approx. 150ng) each of biotin- and digoxigenin-labelled probes

1µl sonicated dog-genomic DNA (10mg/ml)

7.5µl deionised formamide (Fluka)

7.5µl FISH hybridisation buffer (appendix 1)

The tubes were kept in the dark until required. The probe mixtures were denatured at 70°C for ten minutes and then pre-annealed at 37°C for 30 minutes.

The metaphase chromosome spreads were dehydrated in an alcohol series (2 x 70%, 2 x 90%, 1 x 100% ethanol) for three minutes in each solution, air-dried and warmed on the lid of a water bath heated to 70°C. Care was taken to prevent the slides from becoming re-hydrated by exposure to the steam from the water bath. The chromosomes were denatured in 70% formamide (Fluka), 2x SSC (appendix 1) for two minutes at 65°C, then immediately plunged into ice-cold 70% ethanol and left for three minutes. Dehydration was performed through the alcohol series as described above. The slides were air-dried and pre-warmed to 38°C just before use.

The entire volume of denatured probe was placed onto the denatured metaphase spread. A clean 22mm x 22mm coverslip was placed on top and sealed with Cow gum (Cow Proofings Ltd). The slides were incubated at 37°C in a box humidified with 4x SSC for 16 - 18 hours.

2.8.4 Probe detection

The slides were removed from the humidified box and the Cow gum seal and coverslips removed by washing in 2x SSC. The detection protocol was carried out as follows:

1. Slides were placed in a Coplin jar (BDH) containing 50% formamide, 2x SSC at 42°C for three minutes. The jar was sealed with Parafilm (Sigma) and inverted several times during the incubation. This wash was repeated twice more with fresh 50% formamide, 2x SSC at 42°C.
2. Three more three minute washes were performed in 2x SSC at 42°C as described in step 1.
3. The slides were incubated for 30 minutes at 42°C with FISH blocking wash solution (appendix 1).
4. They were removed and 120µl detection layer 1 (FISH detection antibodies, appendix 1) were pipetted over the slides' surface. They were then covered with a piece of Parafilm and placed at 37°C in the humidified box for 30 minutes.
5. The slides were removed and washed in three changes of FISH blocking wash solution at 42°C for three minutes each, with agitation.

6. They were removed and 120µl detection layer 2, (appendix 1) were added as in step 4.
7. The slides were removed and washed as in step 5.
8. The slides were removed and 120µl detection layer 3, (appendix 1) were added as in step 4.
9. The slides were removed and washed in three changes of 2x SSC at room temperature, for three minutes each.

The chromosomes were counterstained by pipetting 100µl 4', 6-diamidino-2-phenylindole (DAPI, 80ng/ml, Vector) over the slide's surface and then incubated at room temperature for five minutes. Surplus DAPI was washed off with 2x SSC and the slides allowed to air dry. Approximately 30µl antifade mountant (Vectashield, Vector) were placed on each slide and covered with a clean 22mm x 50mm coverslip, then compressed against 3MM paper to remove any excess liquid. The slides were examined with a Smart Capture FISH station (Vysis) comprising a fluorescent microscope (Axiophot, Zeiss) equipped with Texas Red, FITC and DAPI filters (Chroma Technologies) and a cooled CCD camera (KAF1400, Photometrics) driven by dedicated software (IPlab Spectrum, Vysis) on a Quadra computer (Apple Macintosh). A minimum of 30 chromosomes were examined for each set of probes, i.e. 15 pairs of chromosome 1 or chromosome X (from a female spread, or 30 chromosome X from a male spread).

2.9 Sequence analysis and primer design

Vector and repeat elements were removed from both the small insert and cosmid sequences by exporting the sequence files to a word-processing document and deleting the unwanted sequence. The remaining sequences were examined using the BLAST algorithm (Altschul *et al.* 1997) against the GenBank database (Benson *et al.* 1999) to identify features of interest.

Oligonucleotide primers were designed to sequences flanking the microsatellites, or to other regions of interest identified by the BLAST alignments, such as regions with similarity to genes in other mammalian species, using the computer programme Primer (version 0.5, Whitehead Institute). Primer sequences of 18 - 26 bases were selected with annealing

temperatures of between 46 and 64°C. The primers were synthesised by commercial suppliers (Genset & Pharmacia) and stored at -20°C.

In addition to the flow-sorted small insert and cosmid libraries, primers for type-1 loci were identified from published work for several X-linked genes. These included comparative anchor tagged sequences (CATS, Lyons *et al.* 1997), universal mammalian sequence-tagged sites (UM-STS, Venta *et al.* 1996) and traced orthologous amplified sequence tags (TOAST, Jiang *et al.* 1998). Suitable primers were synthesised commercially and examined for their suitability to be mapped by linkage and/or radiation hybrid mapping.

2.10 PCR optimisation

Conditions for the PCR (Saiki *et al.* 1986) were optimised with a range of concentrations of $MgCl_2$ (1 - 3mM) and DNA from the clone, genomic dog and hamster cell line A23. Optimised primer sequences and PCR conditions are given in appendix 2

The PCR mixture was prepared as follows:

80ng each of forward and reverse primers

0.2mM dNTPs (Pharmacia)

$MgCl_2$ to give final concentrations of 1, 1.5, 2, 2.5, 3 mM (Perkin Elmer)

1µl 10x PCR buffer 2 (Perkin Elmer, appendix 1)

1 unit AmpliTaq Gold (Perkin Elmer)

1µl DNA (approx. 50ng/µl)

Made up to 10µl with sterile MQ water

Either touchdown (Don *et al.* 1991) or standard PCR (Saiki *et al.* 1986) protocols were used, as described below. The cycling annealing temperature (t_{ann}) was chosen 3°C lower than the calculated melting temperature (provided by the suppliers of the primers, Genset and Pharmacia) of the primers. If this failed to give satisfactory amplification, other annealing temperatures were tried until a clean PCR product was obtained. PCR products were examined by gel electrophoresis (8 V/cm) on a 3% agarose gel (NuSieve, Flowgen) in 1x TAE, alongside 1kb size-standard ladder (Life Technologies). Gels were stained with ethidium bromide (0.5ng/µl) and examined under UV illumination.

Touchdown PCR used at the Animal Health Trust

95°C	25mins	1 cycle[+]
94°C	1min}	
t_{ann} + 9°C to t_{ann}+ 1°C	1min}	18 cycles*
72°C	1min}	
93°C	1min}	
t_{ann}	1min}	16 cycles
72°C	1min}	
72°C	5 mins	1 cycle

"Long" touchdown PCR used at the Animal Health Trust**

95°C	25mins	1 cycle[+]
94°C	1min}	
t_{ann} + 9°C to t_{ann}+ 1°C	1min}	18 cycles*
72°C	2mins}	
94°C	1min}	
t_{ann}	1min}	16 cycles
72°C	2mins}	
72°C	5 mins	1 cycle

Standard PCR used at the Animal Health Trust

95°C	25min	1 cycle[+]
94°C	1min}	
t_{ann}	1min}	30 cycles
72°C	1min}	
72°C.	5min	1 cycle

*Notes: *Two cycles each are performed with annealing temperatures of t_{ann} + 9°C, t_{ann} + 8°C, t_{ann} + 7°C ... t_{ann} + 1°C, where t_{ann} is the annealing temperature of the primers, calculated as described above.*

[+]It was found that AmpliTaq Gold, required a longer activation step than recommended in the manufacturer's instructions; 25 minutes were determined empirically to be a robust activation period. When normal Taq DNA polymerase (which does not require activation) was used for PCR, this step was removed from the thermocycler's programme.

***This programme was used for larger (>500bp) or for products difficult to amplify with the usual method.*

2.11 Linkage analysis

Polymorphic microsatellites markers may be used to generate genetic linkage maps using reference families of a suitable structure. In this study, microsatellite markers (sections 2.3 & 2.9 above) were tested on a limited number of individuals to determine whether they were

polymorphic in the family. Those that proved to be polymorphic were then genotyped against the rest of the reference family.

2.11.1 Assessment of heterozygosity

Typically, several markers were assessed for polymorphism on each polyacrylamide gel. The fluorescent rhodamine dyes, R6G and R110 (Perkin Elmer) are supplied attached to 2'-deoxyuridine 5'-triphosphate (dUTP) and are incorporated into the extending DNA strand during the PCR. These two dyes have different excitation properties, which allowed PCR products from several markers to be run in each lane of the gel. However, as these dyes have some spectral overlap, markers were selected so that there was at least 10bp difference in the sizes of the PCR products, to prevent bands overlapping on the gels and thus the signals interfering.

PCRs were prepared as follows:

50ng each of forward and reverse primers

0.2mM dNTPs (Pharmacia)

50pmoles of either R6G or R110, (Perkin Elmer)

PCR buffer 1 (to give 1.5mM $MgCl_2$ final conc.; Perkin Elmer) or PCR buffer 2 (Perkin Elmer) and 25mM $MgCl_2$ to obtain the appropriate concentration (Perkin Elmer)

1 unit AmpliTaq Gold (Perkin Elmer)

Approx. 50ng genomic DNA

Total volume made up to 10µl with MQ water.

The PCR was carried out using a thermal cycler (Tetrad, MJ Technologies) under the optimised conditions for each marker as shown in appendix 2. PCR products were precipitated by the addition of $^1/_{10}$ volume of 3M sodium acetate (appendix 1) and 2.5 volumes of ice-cold absolute ethanol. The products were pelleted by centrifugation at 10,000g (Eppendorf) at 4°C for 30 minutes, the supernatant removed, each pellet was washed with 500µl ice cold 80% ethanol, respun as before and air-dried

Denaturing polyacrylamide gels were prepared to give a final concentration of 4.25% acrylamide according to the manufacturer's instructions (Anachem). The dried sequencing products were prepared for running on the gel by the addition of 3.5µl sequencing gel loading buffer (appendix 1) and 0.5µl fluorescently labelled size standard, typically ROX500 (Perkin Elmer) (Mayrand *et al.* 1992). The samples were denatured at 95°C for two minutes then placed

on ice and 2μl of each sample were loaded onto the gel. The PCR products of three markers were run in one lane with the internal size standards. Gel electrophoresis was carried out for two - three hours according to the manufacturer's recommendations so that at least two size standards beyond the largest microsatellite band had been electrophoresed past the laser (this ensured accurate estimation of fragment sizes).

Analysis was carried out using the manufacturer's software (Genescan 2.0.2 and Genotyper 1.1.1, Perkin Elmer) and alleles scored by the software relative to the internal size standards run with the samples.

2.11.2 Genotyping of pedigrees

The protocol described above was used to genotype the DNA from members of the reference family (Mellersh *et al.* 1997), where the typing on the parents showed that the family was informative for that microsatellite.

2.11.3 Analysis

Analysis of the genotyping results was carried out using the manufacturer's software (Genescan 2.0.2 and Genotyper 1.1.1, Perkin Elmer). The allele sizes were estimated by the software by comparison to the internal size standards run with the samples. The results were manually entered into a spreadsheet (Excel, Microsoft) then exported into a word-processing programme (Notepad, Microsoft) for input into the analysis software. Linkage analysis was performed using the MLINK option of the Linkage programme (Lathrop *et al.* 1984). CRI-MAP (Lander & Green, 1987) was used to integrate the data with the existing map. Lod scores greater than 2.0 and 3.0 were considered valid indications of linkage for chromosomes X and 1 respectively (Ott, 1991).

2.12 Radiation hybrid analysis

Primers designed from clones from the chromosome-specific small-insert libraries were used with the dog whole-genome radiation hybrid panel, T72 (Research Genetics) to map markers. The primers used were those that flanked microsatellites, STS and other regions of interest obtained from BLAST alignments, such as regions with similarity to genes in other mammalian species. In addition, primers from comparative anchor tagged sequences (CATS,

Lyons *et al.* 1997), universal mammalian sequence-tagged sites (UM-STS, Venta *et al.* 1996) and traced orthologous amplified sequence tags (TOAST, Jiang *et al.* 1998) known to map to human and mouse chromosome X were also tested.

2.12.1 Radiation hybrid typing protocols

2.12.1.1 Standard method

The 96 members of the dog whole-genome radiation hybrid panel (including controls of two each of hamster and dog-genomic DNA, and one of water; Research Genetics) were diluted to 15ng/μl and dispensed into 96-well plates. The plates were sealed with adhesive film (Alpha Laboratories) and stored at –20°C

Typing reactions (10μl) consisted of

80ng each of forward and reverse primers

0.2mM dNTPs (Pharmacia)

$MgCl_2$ (at optimised conc. section 2.10; Perkin Elmer)

1x PCR buffer 2 (Perkin Elmer; appendix 1)

1 unit AmpliTaq Gold (Perkin Elmer)

45ng panel DNA

Reactions were prepared as a master mix containing all components except the panel DNA, dispensed into either 0.2ml strip tubes (Robbins Scientific) or microtitre plates (sealed for cycling with adhesive film; Alpha Laboratories) and then panel DNA was added. The PCR was carried out using a thermal cycler (Tetrad, MJ Technologies) under the optimised conditions for each marker (see appendix 2). Agarose gel loading buffer (appendix 1) was added (5μl per sample) and PCR products analysed on 4% agarose gels in 1x TAE (appendix 1) at 4V/cm for approximately one hour.

Subsequently, the gels were stained with 0.05mg/ml ethidium bromide for 60 - 90 minutes and examined with an UV transilluminator as described previously. Gels were photographed under UV illumination and scored for the presence or absence of a PCR product for each radiation hybrid or control DNA for each marker. Hard copies of the gel image were obtained (Mitsubishi video processor) and the results recorded in a spreadsheet (Excel, Microsoft). All the reactions were carried out in duplicate so any inconsistencies could be observed, and if necessary, the PCR repeated for a third time.

2.12.1.2 Additional RH typing method from the Sanger Centre

A number of markers were typed using a different protocol, used in the human genome radiation hybrid mapping project at the Sanger Centre, Cambridge UK. This method has been successfully used in the generation of the second gene map of the human genome (Deloukas *et al.* 1998). It was used in this project to investigate whether it was better able to specifically amplify the dog fraction of the hybrids.

Optimisation reactions (20µl) consisted of:

12.5% sucrose

1x Sanger PCR buffer (appendix 1)

0.187µl reducing solution (appendix 1)

0.2mM dNTPs (Pharmacia; appendix 1)

0.6U AmpliTaq R (Perkin Elmer)

50ng template DNA

25ng each of forward and reverse primers

PCR conditions are of the touchdown type (Don *et al.* 1991) as follows:

94°C	5 mins	1 cycle
93°C t_{ann} +5°C to t_{ann} 72°C	30 secs} 50 secs} 50 secs}	10 cycles$^{\diamond}$
93°C t_{anneal} 72°C	30 secs} 50 secs} 50 secs}	30 cycles
72°C	5 mins	1 cycle

$^{\diamond}$ *one cycle is performed at each of* t_{ann} *+5°C,* t_{ann} *+4.5°C,* t_{ann} *+4°C ...* t_{ann}.

In this protocol, the optimal annealing temperature was established empirically using hamster cell line A23, dog genomic and pooled WG-RH T72 DNA. Temperatures (t_{ann}) of 50, 55, and 65°C, were tested and the optimum temperature used for typing the panel.

Reactions were carried out in 0.2ml thermocycler plates (Costar) and the plates sealed using reusable sealing mats (Hybaid). These were prepared for re-use by exposing to ultra-violet light in a Stratalinker (Stratagene) at maximum energy (240mJoules) twice. The reducing solution contained β-Mercaptoethanol, which is highly volatile and so was diluted immediately prior to use. All reagents were kept on ice. The presence of sucrose in the reaction mix allowed

the PCR products to be loaded directly onto an agarose gel. Gels were used at 2.5% agarose (Gibco-BRL) in 1x TBE (appendix 1) containing 50ng/μl ethidium bromide (Sigma). Gel electrophoresis was carried at 9 V/cm for 20 minutes in 1x TBE containing 0.5μg/ml ethidium bromide. Gels were examined under UV light and a permanent record made by taking a Polaroid photograph of the gel.

2.12.2 Radiation hybrid analysis

The typing data was manually entered into a spreadsheet (Excel, Microsoft) and exported into a word-processing programme (Notepad, Microsoft) for input into the analysis software. The results were initially analysed using the RH2PT option of the computer programme RHMAP version 3.00 (Boehnke *et al.* 1991) to determine lod scores and linkage between markers. Subsequently, the order of markers within the linkage groups was determined using the MINBRK and MAXLIK options of the same software.

Appendix 2 lists the conditions used for each of the markers used in the radiation hybrid mapping.

3 Results

3.1 Flow-sorting of chromosomes for library construction

Chromosomes X and 1 are the two largest chromosomes and are well separated from the others in the dog flow karyogram, shown in figure 5. Flow-sorted libraries constructed by the Sanger centre had been found to require a minimum of 250,000 chromosomes (M. Ross, personal communication). Dog chromosomes need to be sorted in one session as the chromosome preparations degrade rapidly and become unsuitable for sorting (C. Langford personal communication). This limited the material sorted to approximately 200,000 events (chromosomes) for each of chromosomes X and 1. Although this was less than had been used previously to make similar flow-sorted libraries, it was considered sufficient to try to construct the two small-insert libraries. The limited amount of material meant that it was decided not to test the purity of the flow-sort directly by a cytogenetic approach, where a small number of sorted chromosomes (approximately 5,000) are fixed to a microscope slide and an aliquot of the appropriate chromosome paint is hybridised to them. In this study, the purity of the libraries could be investigated by the numbers of clones that hybridised to the expected chromosomes in FISH studies.

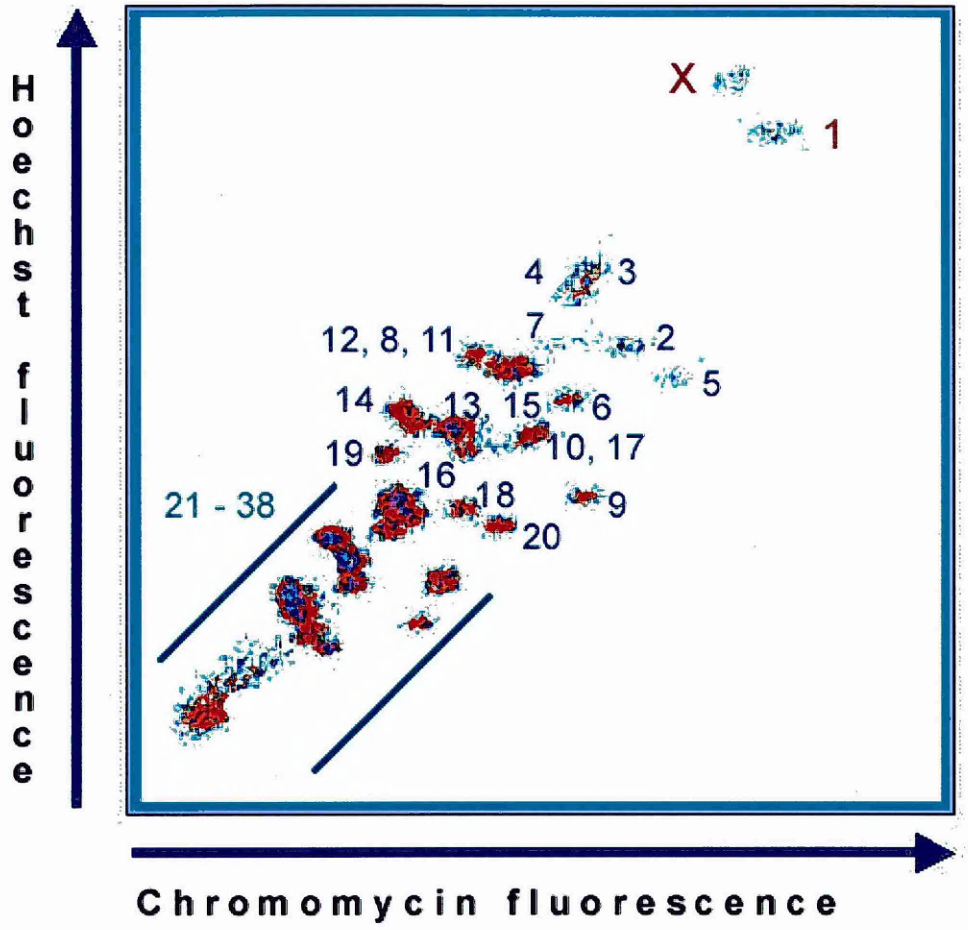


Figure 5 Bivariate flow karyotype of a female dog

The two largest chromosomes, X and 1, were sorted and used to construct two chromosome-enriched, small insert libraries. Hoechst 33258 has AT base pair specificity and chromomycin A3 has GC base pair specificity. Hence, X and 1 are the largest chromosomes, X being more AT-rich than 1. The numbering of each peak is according to Langford et al. (1996)

3.2 Small insert libraries - preparation

The chromosome-enriched libraries were plated onto media containing IPTG and X-gal as described in section 2.2. The colonies containing insert DNA from the flow-sorted chromosomes appeared white under the selection system used whereas non-recombinants were blue. The numbers of blue and white colonies were noted for each library and the percentage of recombinants calculated. The yields (recombinants per microgram of insert DNA) were calculated. These results are shown below in tables 2 and 3.

3.2.1 Chromosome X

Table 2 Recombinant and non-recombinant colonies from chromosome X library

	Chromosome X
Total numbers of colonies	7730
Recombinant (white) colonies	6048
Non-recombinant (blue) colonies	1682
Recombinant colonies %	78%
Yield (recombinants/μg insert DNA)	3×10^6

3.2.2 Chromosome 1

Table 3 Recombinant and non-recombinant colonies from chromosome 1 library

	Chromosome 1
Total numbers of colonies	5923
Recombinant (white) colonies	4416
Non-recombinant (blue) colonies	1507
Recombinant colonies %	75%
Yield (recombinants/μg insert DNA)	2.2×10^6

There is little comparative data for this method of library construction. However, it may be assumed that the integrity of the DNA is affected by the processing involved in flow sorting, which probably has a detrimental effect on the transformation efficiency. Nonetheless, there were sufficient recombinant clones to screen for microsatellites given that the estimated frequency of dinucleotide repeats in the dog genome is every 42kb (Rothuizen *et al.* 1994). The

yield was lower for the chromosome 1 library and consequently fewer colonies were obtained for this library than for the X library; however, the percentage of recombinant clones was similar. It is not apparent why the chromosome 1 library yielded fewer colonies.

3.3 Small insert libraries - screening for microsatellites

An example of the results of screening the chromosome-enriched libraries with α-^{32}P-dATP labelled poly- $(dA\text{-}dC)_n$. $(dG\text{-}dT)_n$ probe is shown in figures 6a and 6b where the colonies were replicated by hand using a 96-well format. Figure 6a shows the autoradiograph produced after the low stringency 42°C wash with the putative positive colonies visible above the background signal. Figure 6b shows the autoradiograph after the higher stringency wash at 65°C when only two positive signals remain. At this stringency, only inserts containing dinucleotide repeats $(CA)_n$, where $n \geq 12$, are expected to give a signal. Hence, the other colonies that appeared positive in figure 6a may contain shorter and/or imperfect repeats. Figure 7a shows the autoradiograph of colonies replica plated in 384-well format following a low stringency wash and figure 7b, the autoradiograph after the high stringency wash. Comparison of the autoradiographs from both washes allowed the position of the positive colonies to be determined. The correct identification of the positives was particularly important with the small, closely spaced colonies produced from replica plating from the 384 well plates. The positive clones were picked and allocated a serial number, prefixed with the an "x" for the X-chromosome library and "1-" for the chromosome 1 library, for example, x24 or 1-30.

3.3.1 Chromosome X

Ninety-five clones from the chromosome X library were positive for the poly (CA) probe after high stringency washes, representing 1.6% of the clones from this library.

3.3.2 Chromosome 1

Sixty clones from the chromosome 1 library were positive for the poly (CA) probe after high stringency washes, representing 1.4% of the clones from this library.

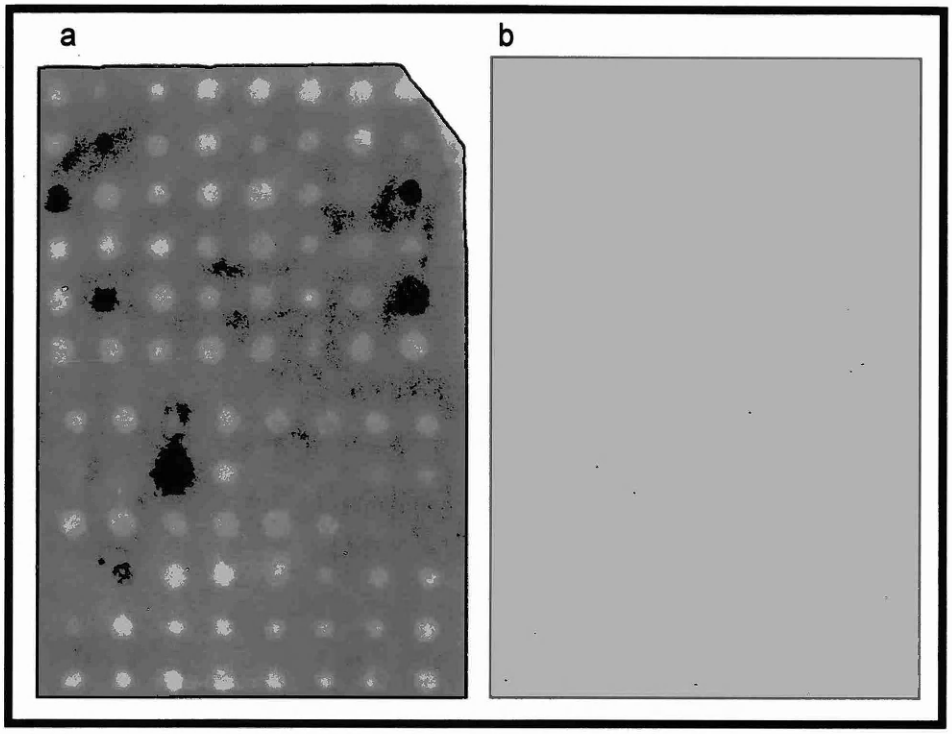


Figure 6 screening of small insert clones for microsatellites.
Clones were replica plated in a 96-well format. (a) autoradiograph after low-stringency wash and (b) after high stringency wash

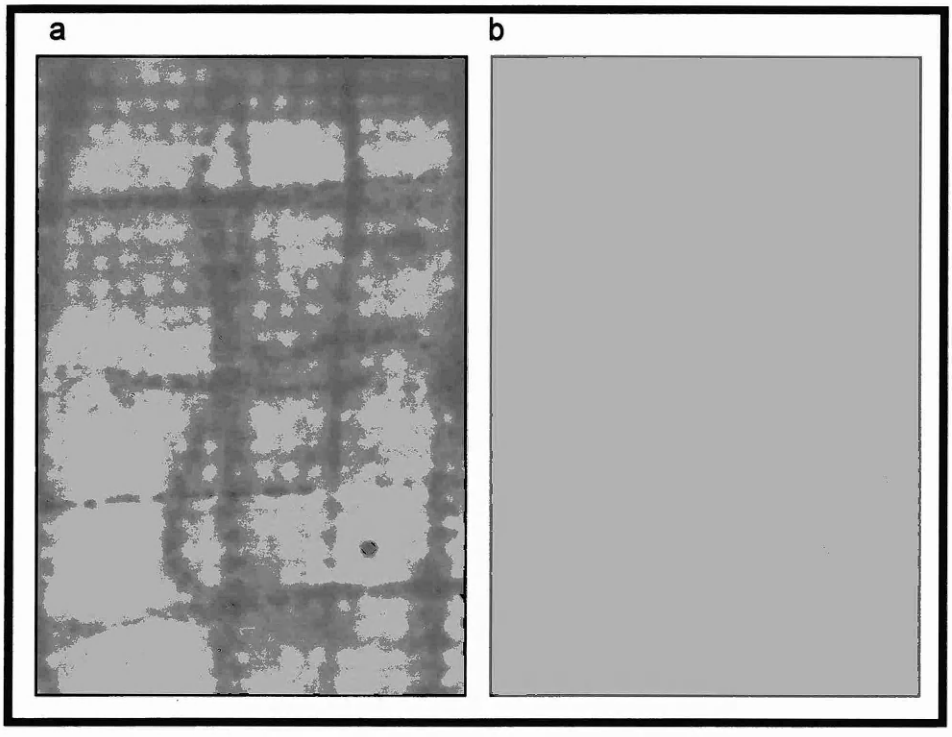


Figure 7 screening of small insert clones for microsatellites.
Clones were replica plated in a 384-well format. (a) autoradiograph after low-stringency wash and (b) after high stringency wash

3.4 Small insert clone sequencing

Positive clones were cultured and plasmid DNA prepared and quantified as described in section 2.3, yielding an average of 480ng plasmid from each 1ml culture of each clone. Examples of typical sequencing traces are shown in figures 8a, b, c and d, showing T3, T7, DRP and back-sequencing reactions respectively.

Sequencing reactions using T3 and T7 primers generally gave higher quality results with longer readable sequences than the repeat primers. Typically, at least 400bp of sequence was obtained with T3 and T7 primers. The sequencing reactions using the DRPs are dependent on the sequence of the repeat being uninterrupted for length of the primer, and are more prone to mis-priming. However, there was generally sufficient sequence from the DRP primers to design a primer to back-sequence through the repeat. Sequencing through the microsatellite was straightforward although the presence of long polyA regions was found to inhibit the sequencing reactions.

Sequences from the confirmed chromosome X- and 1-derived small-insert and cosmid clones (as determined in later parts of this study) have been submitted to EMBL (URL: http://www.ebi.ac.uk/), after the removal of any vector sequence, and are listed below:

Clone name	Accession no.	Clone name	Accession no.	Clone name	Accession no.
AHTX1	AJ403670	AHTX64	AJ403700	AHTH223	AJ405176
	AJ403671		AJ403701	AHTH185	AJ405177
AHTX13	AJ403668		AJ405151	AHTH152	AJ405216
	AJ403669	AHTX66	AJ403702	AHT1-01	AJ405194
AHTX21	AJ403672		AJ403703	AHT1-03	AJ405195
	AJ403673	AHTX67	AJ403704	AHT1-09	AJ405169
AHTX24	AJ403674		AJ403705	AHT1-11	AJ405196
	AJ403675	AHTX68	AJ405153	AHT1-20	AJ405197
AHTX27	AJ403676		AJ405162		AJ405198
	AJ403677	AHTX69	AJ405154	AHT1-21	AJ405199
AHTX30	AJ403678		AJ405163		AJ405170
AHTX31	AJ403679	AHTX7	AJ405155	AHT1-27	AJ405200
AHTX33	AJ403680		AJ405164	AHT1-30	AJ405201
AHTX34	AJ403681	AHTX71	AJ405184		AJ405202
AHTX35	AJ403682		AJ405185	AHT1-32	AJ405203
	AJ403683	AHTX73	AJ405179	AHT1-33	AJ405171
AHTX36	AJ403684		AJ405183	AHT1-34	AJ405204
	AJ403685	AHTX74	AJ405180		AJ405205
AHTX38	AJ403686		AJ405152	AHT1-42	AJ405206
	AJ403687	AHTX78	AJ405181		AJ405172
AHTX4	AJ403693		AJ405189	AHT1-43	AJ405207
AHTX40	AJ403688	AHTX9	AJ405158	AHT1-45	AJ405208
	AJ403689		AJ405159	AHT1-46	AJ405209
AHTX47	AJ403690	AHTX91	AJ405182	AHT1-49	AJ405173
AHTX49	AJ403691		AJ405156	AHT1-57	AJ405210
	AJ403692	AHTX92	AJ405178	AHT1-57	AJ405174
AHTX50	AJ403694		AJ405157	AHT1-58	AJ405211
AHTX56	AJ403695	AHTH38	AJ405212	AHT1-60	AJ405175
AHTX57	AJ403696	AHTH32	AJ405213		
AHTX58	AJ403697	AHTH304	AJ405214		
	AJ403698	AHTH254	AJ405215		

3.4.1 Chromosome X

Sequencing the putative microsatellite-containing clones from the small insert library with T3 and T7 primers revealed that nine of the 95 clones did not contain an insert. These non-insert clones were all picked robotically and could have originated from colonies with a blue periphery and white centre. Such colonies, which were observed but rejected in manual pickings, are presumed to be unstable and may have lost their microsatellite-containing inserts during plasmid preparation. Of the 86 insert-containing clones, the T3 /T7 sequencing revealed microsatellites in 23. The remaining 63 clones were sequenced using the DRPs and back sequencing with unique primers. This yielded a further 37 microsatellites. The remaining 26 sequences may have contained a microsatellite, but failed to sequence adequately, presumably due to the presence of compound or interrupted repeats. Of the 60 microsatellites identified, three were unsuitable for mapping since they were located adjacent to the vector sequence and ten were unsuitable due to their proximity to other repetitive elements such as SINEs or LINEs. Thus, from the initial 95 clones, 47 were suitable for use in microsatellite-dependent mapping. These results are summarised in table 4.

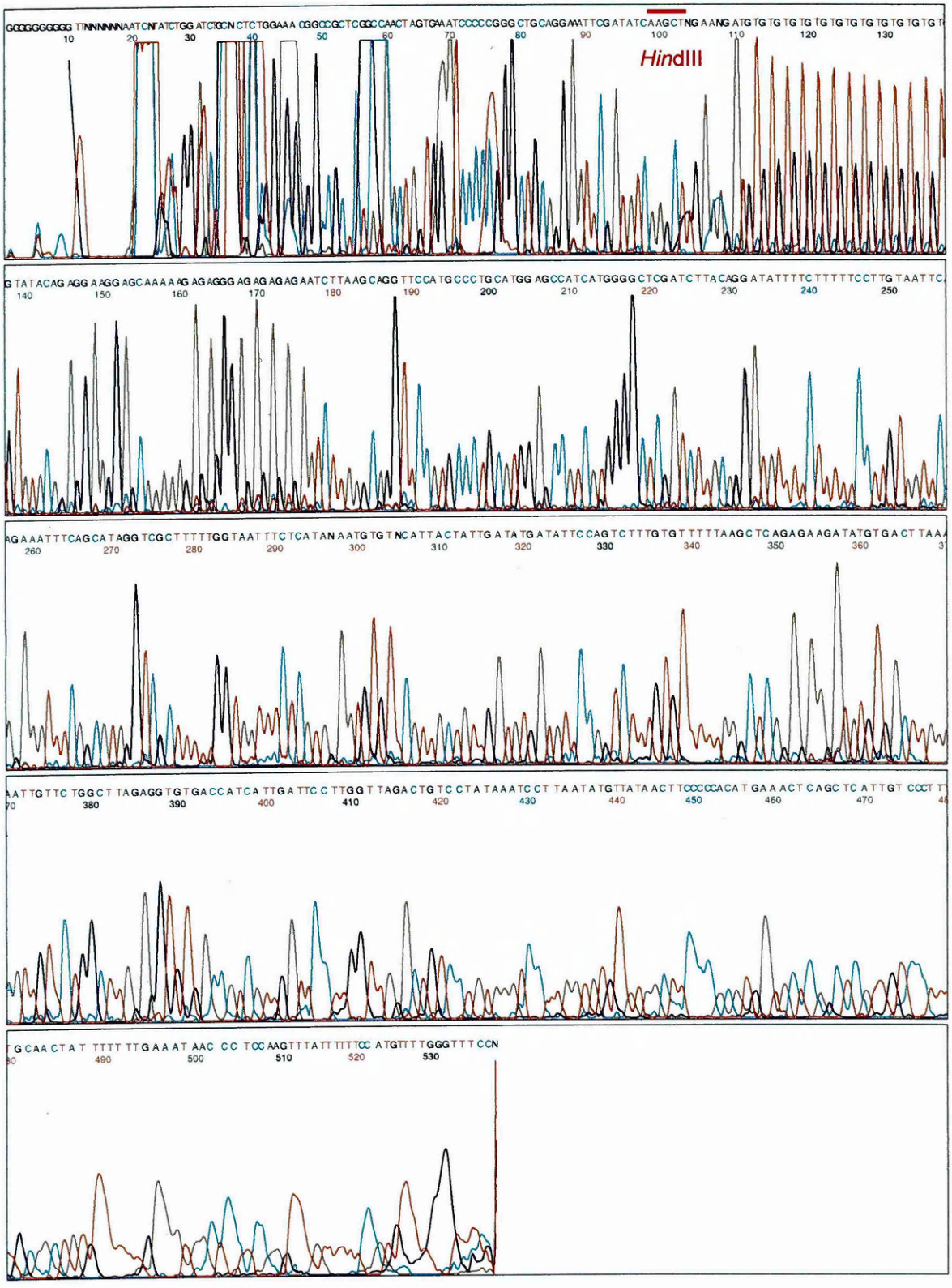


Figure 8a Electropherogram from clone x75 sequenced with T3 primer.
In this instance, the microsatellite (TG_{14}) is located too close to the HindIII cloning site (red bar) to allow a flanking primer to be designed on one side.

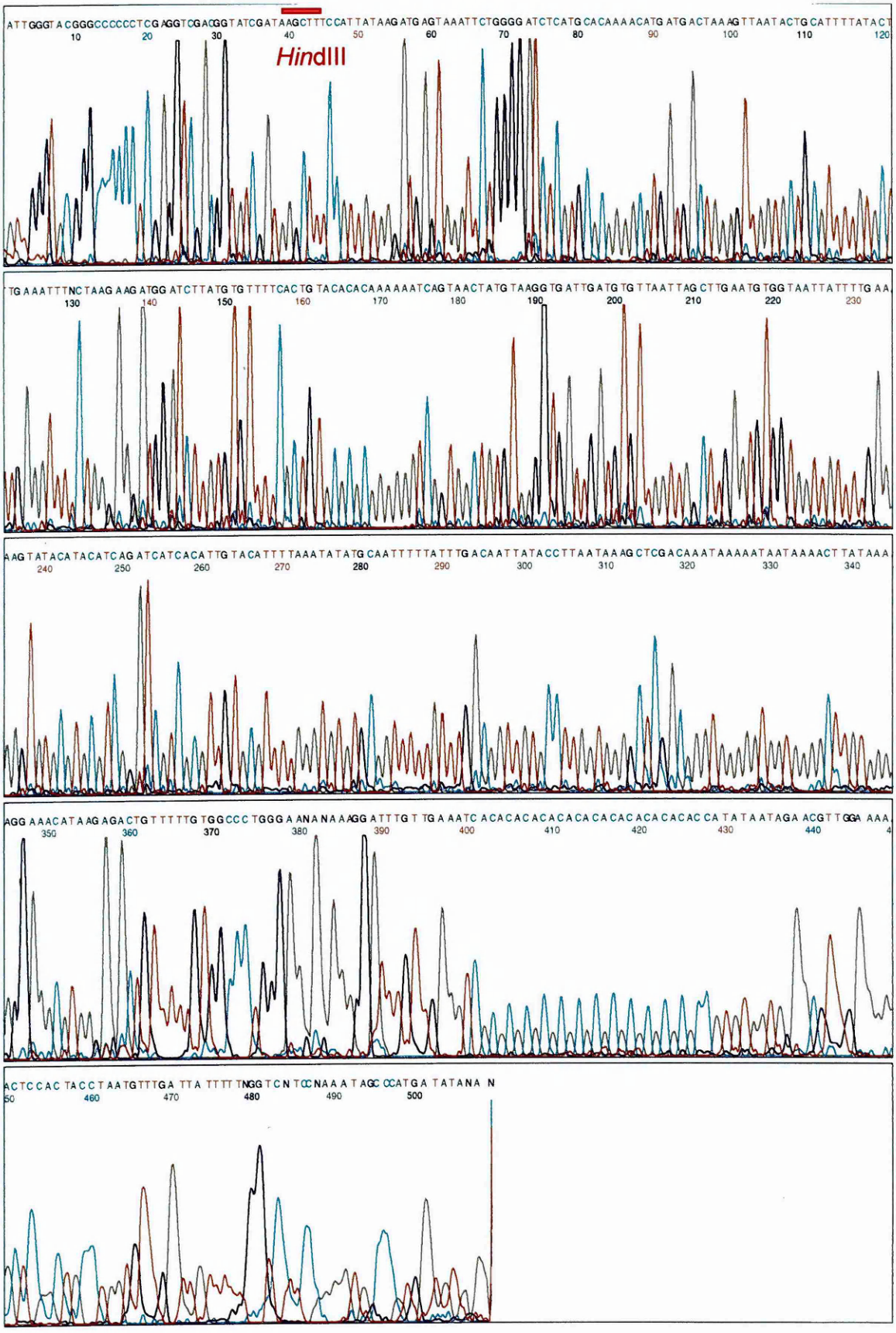


Figure 8b Electropherogram from clone x68 sequenced with T7 primer.
A microsatellite (CA_{13}) is located at bases 401 - 426, primers were designed to flank the repeat. The HindIII site is marked by the red bar, the sequence to the left of it is from the vector, and to the right, the insert sequence.

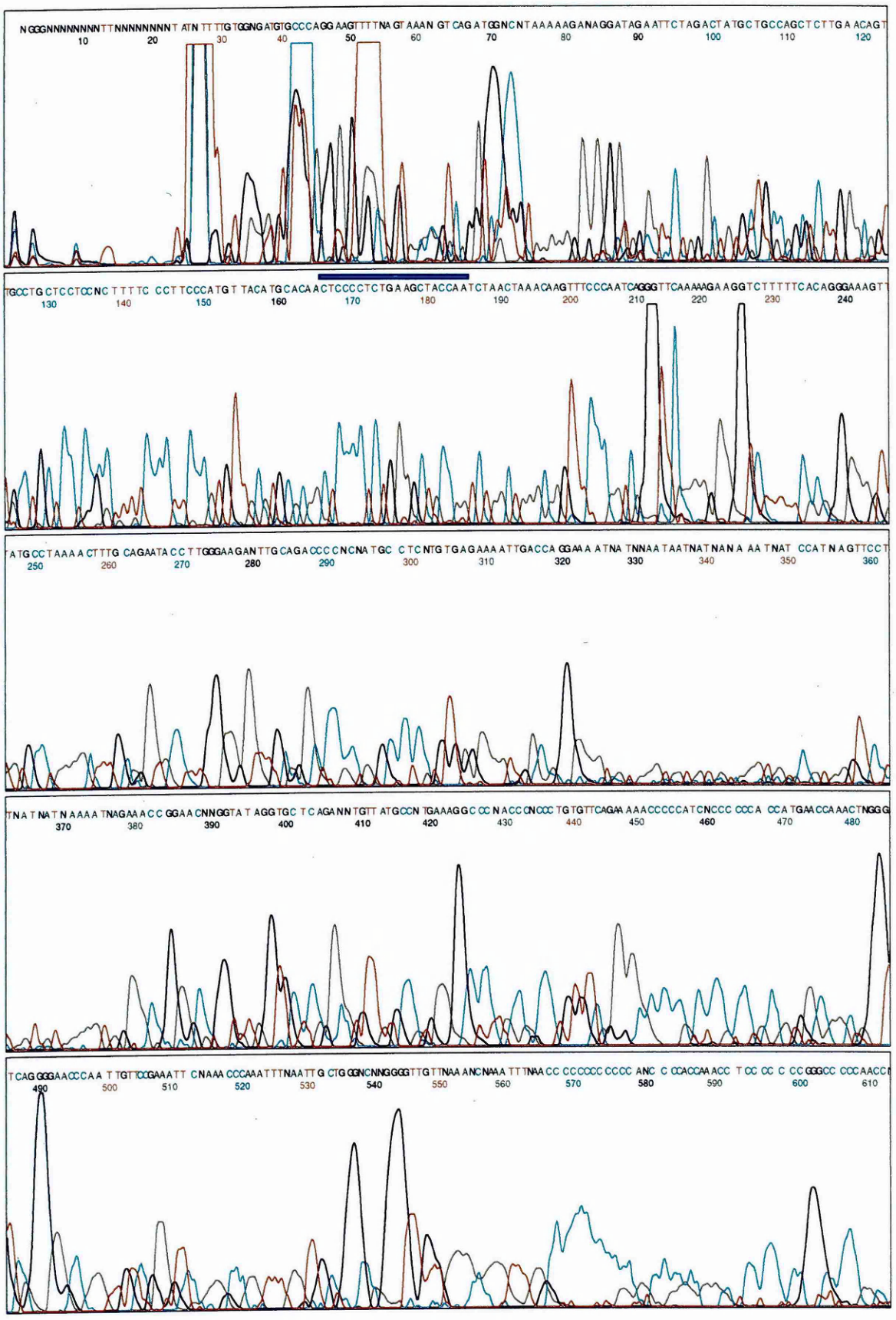


Figure 8c Electropherogram from clone x57 sequenced with DRP $(GT)_{10}A$.
T3 & T7 sequencing did not reveal a microsatellite, so this clone was sequenced with all 6 DRPs. The primer $(GT)_{10}A$ produced readable sequence and a unique primer (marked by blue bar) was designed to sequence back through the repeat.

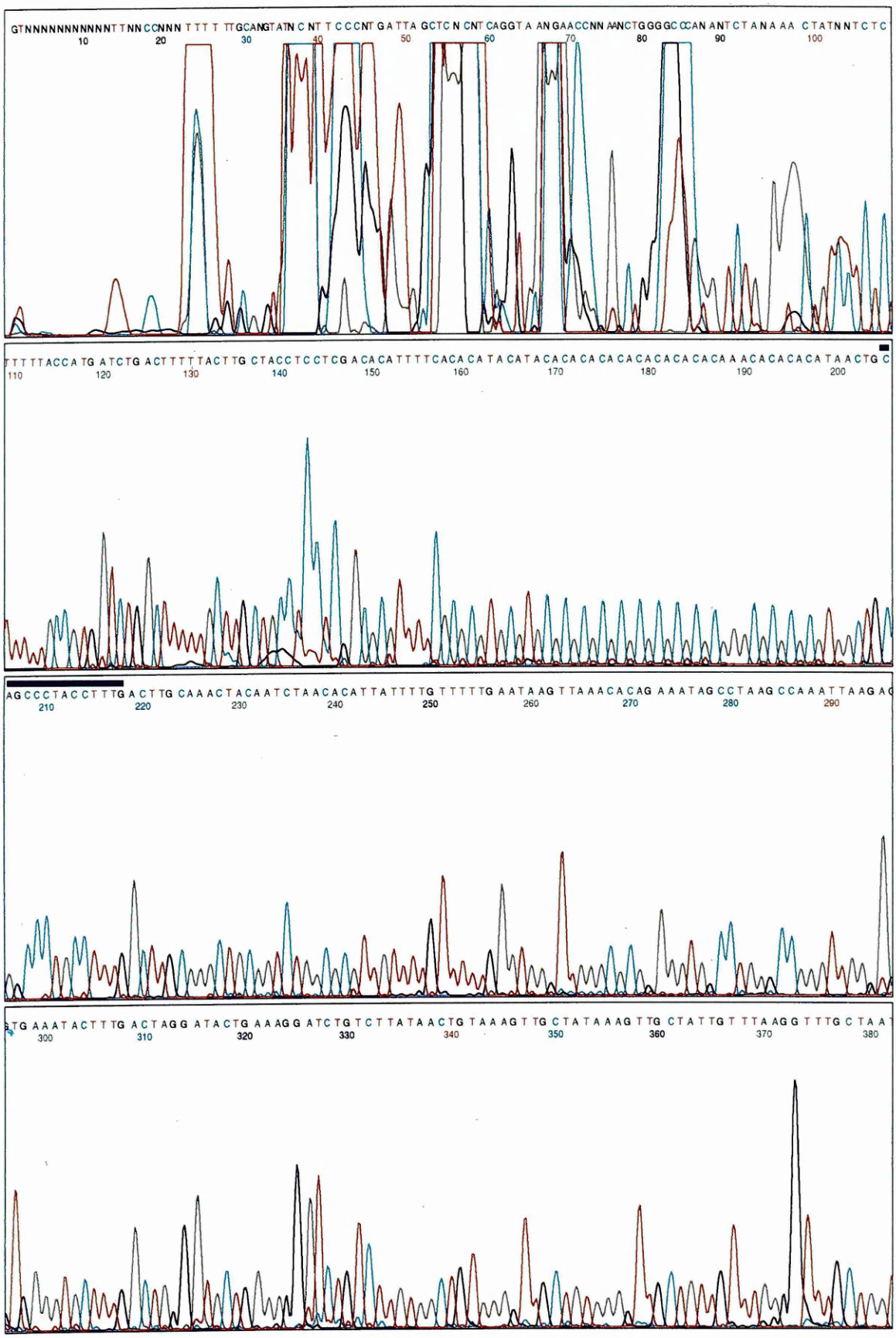


Figure 8d Electropherogram from x57 with the unique primer designed from DRP sequencing. *The microsatellite (AC_{10}) was revealed and the companion flanking primer designed (marked by green bar). In this particular case, the unique sequencing primer was able to be used as one of the pair of flanking primers.*

3.4.2 Chromosome 1

Sequencing data for the 60 putative microsatellite-containing clones were determined only from the T3 and T7 primers. Four of the clones (picked by the BioPick robot) were found not to contain an insert. Two of the clones (1-28 and 1-29) were found to be identical, presumably due to a picking error and were therefore counted as a single clone. Of the 55 unique inserts, 15 were found by T3 and T7 sequencing to contain a microsatellite, two of which were not usable due to the unavailability of sequence suitable to design primers flanking the repeat (because of proximity to vector or to other repetitive elements). These results are summarised in table 4. The remaining 43 clones were not sequenced with the DRPs but based on the results in 3.4.1, more than half are expected to contain microsatellites.

Table 4 summary of sequencing results from chromosome-enriched libraries

	Chromosome X	Chromosome 1
No. of clones sequenced	95	60
No. of clones without an insert	9	4
No. with microsatellite found by T3/T7 sequencing	23	15
No. with microsatellite found by DRP sequencing	37	Not done
No. of potentially usable microsatellite markers	47	13
No. of clones with at least 1 microsatellite	60	>15

3.5 Cosmid clones - end sequencing

Additional STSs from six cosmid clones known to map to chromosomes X and 1 by FISH (M. Breen, personal communication) were obtained by T3 end sequencing, as described in section 2.7. Sequencing the cosmids with the BigDye chemistry produced usable sequence, with read lengths of typically 400bp, i.e. similar to that obtained from standard dRhodamine chemistry on the small insert clones. An example of a typical sequencing trace from one of the cosmid clones, H32, is shown in figure 9.

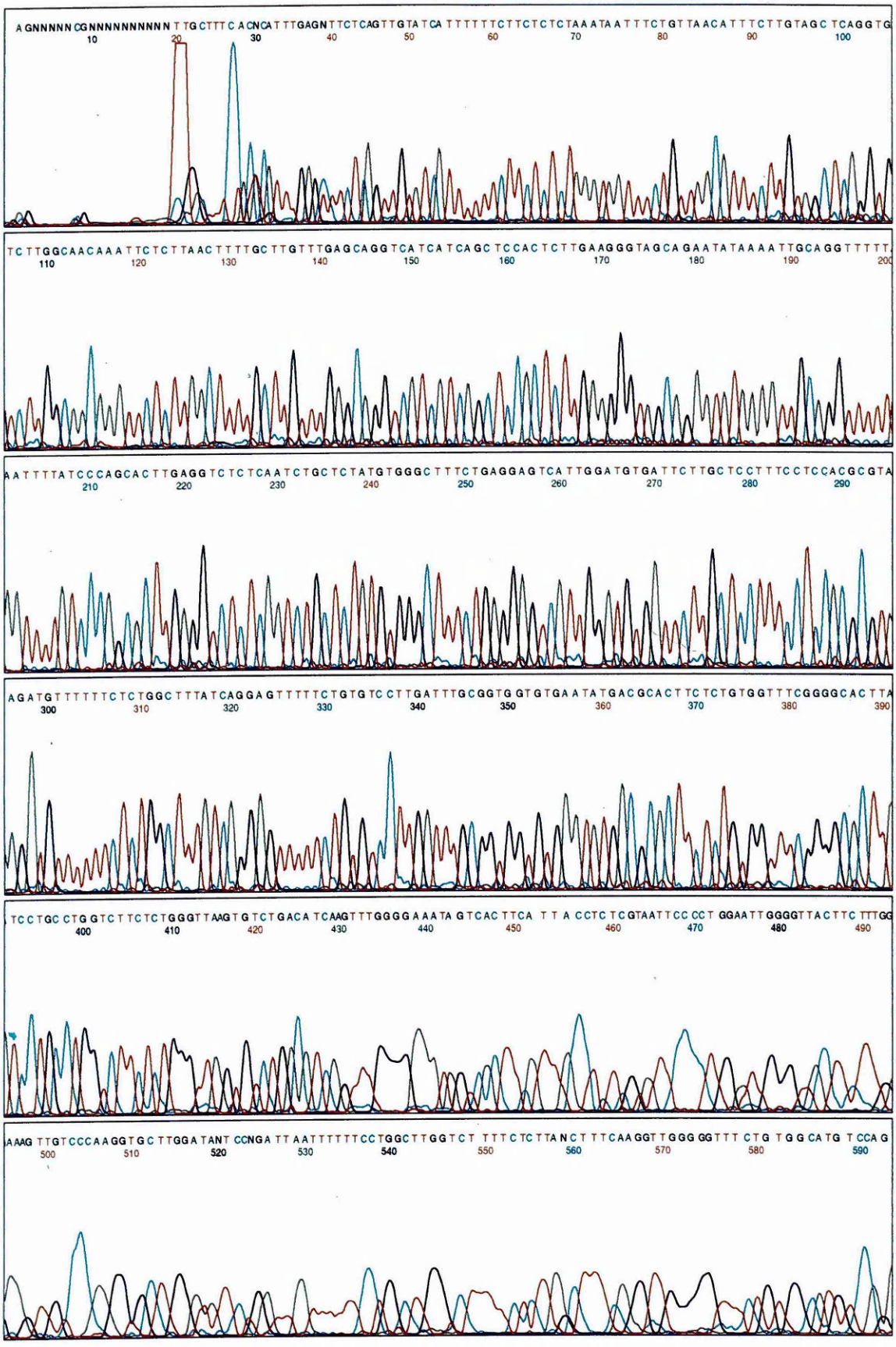


Figure 9 Electropherogram from cosmid H32 sequenced with T3 primer using BigDye chemistry
good quality sequence was obtained up to 450bp.

3.6 Small insert libraries - determination of insert size

Clones containing microsatellites were digested with *Hin*dIII as described previously, to investigate the average insert size of the microsatellite-containing clones produced during library construction. An example of a gel showing the digested fragments from both chromosome X and 1 library clones is given in figure 10. Of the clones analysed, restriction digestion of two, x71 and 1-18, produced three bands, one from the vector and two from the insert. Work carried out at the Sanger centre making human flow-sorted chromosome-enriched small insert libraries, gave an average insert size similar to that obtained in constructing the dog libraries (M. Ross personal communication). The insert size determined from the microsatellite-containing clones may be atypical of the library, since they represent only approximately 1.5% of the flow-sorted chromosome-enriched clones.

3.6.1 Chromosome X

The average size for the microsatellite-containing clones from the X chromosome library is estimated to be 7.5kb with a range from 0.7kb to 12kb. This was determined from those clones that gave one non-vector band on restriction digestion.

In addition to the vector band of 2.9kb, x71 produced a band of approximately 6kb and another, weaker band of approximately 2kb. Since this lower band was much weaker than the others, it is likely that this is due to the presence of a contaminating clone, rather than due to a *Hin*dIII site within the insert (which should yield equimolar fragments). Furthermore, the vector band is brighter for this clone, than any of the others, which would occur if two clones were present simultaneously.

3.6.2 Chromosome 1

The average size for the microsatellite-containing clones from the chromosome 1 library determined from those clones that gave a single *Hin*dIII insert fragment is approximately 6kb, ranging from 0.45kb to 12kb.

One of the clones examined, 1-18, produced three bands, one of the vector, another band of approximately 8kb and one of approximately 5kb. This may be caused by incomplete digestion of chromosomal DNA or co-ligation during library construction, or possibly (since the top band is not as bright as the middle, 5kb band) be due to incomplete digestion of the

plasmid. The two smaller fragments, approximately 5kb and, 2.9kb approximately equal the size of this band, around 8kb.

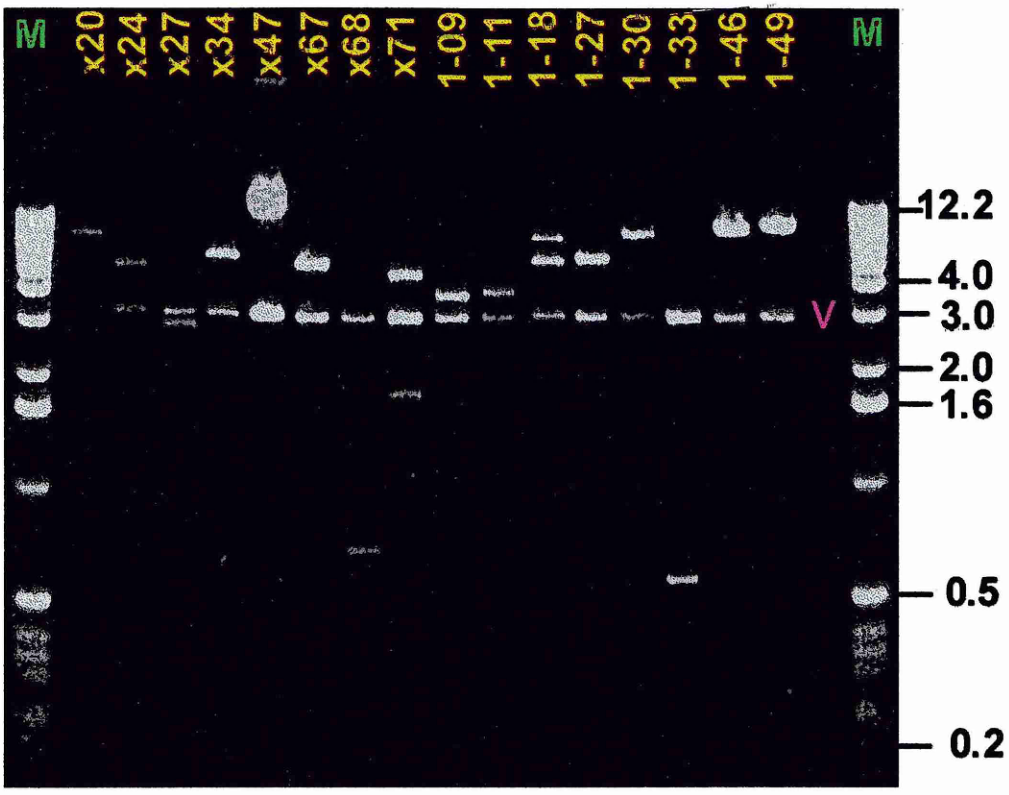


Figure 10 Insert size analysis of 16 microsatellite-containing plasmid clones

X20 to x71: HindIII-digested plasmid DNA from clones of the X-enriched library; 1-09 to 1-49: HindIII-digested plasmid DNA from clones of the 1-enriched library. M indicates the DNA ladder (sizes in kb at right). Vector (v) is approximately 2.9kb.

3.7 Sequence analysis

All the sequences generated in this study were examined for similarity to entries in the GenBank databases (Benson *et al.* 1998) using the BLAST programmes (Altschul *et al.* 1997). BLAST identifies statistically significant similarities between sequences in the database and the query sequence. The significance of each match is reflected by the value, E, which represents the number of database sequences expected, by chance alone, to match the query sequence as well or better than the observed match. An E value of 1 is of very low significance (one such match by chance alone is expected), whereas a value of 10^{-20} is highly significant (the probability of such a match to the total database by chance alone being only 10^{-20}). E values, however, need to be interpreted with caution, and each apparently significant match needs to be examined. For example, apparently highly significant matches between a query sequence and many unrelated database sequences are likely to be due to shared repeat elements (SINE, LINE, MER, microsatellites etc). Conversely, a match with homologous gene sequences from several other species is more likely to be significant. The length of the match is also important: a long contiguous stretch showing moderate similarity is more likely to be relevant than a short or highly fragmented region of higher similarity. (The sequences from this study, after removal of vector, microsatellites and apparent repeat motifs, were typically 200 - 250bp). An example of a significant similarity is the dog sequence from clone x24, which matched database entries from mouse, man and rabbit sequences for phosphorylase kinase alpha-1 (PHKA1), with E values of less than 8.7×10^{-26}; in each case the match extended over 91 - 97% of the dog sequence.

All databases matches returning an E value of $<10^{-5}$ were examined initially. Obvious repeat motifs (SINE, microsatellite etc) were then manually removed from the query sequences, and the search repeated where necessary. In some cases, the expansion of the GenBank database over the course of this study revealed repeat motifs that were not present in initial searches. For example, sequence x12 (from this study) was initially found to contain a SINE element, which was therefore avoided in designing PCR primers for genotyping. Linkage analysis, however, produced results atypical for a sex-linked marker. The reason for this became apparent when later searches against the expanding database revealed a match between x12 and a long terminal repeat mammalian retrotransposon-like superfamily (LTR

MaLR) sequence. Similarly, a LINE1 element, recently added to GenBank, was found to be present in clone x66 (from this study) in a region to which PCR primers had already been designed. These primers had also given results inconsistent with a typical sex-linked marker. An overview of the database searches for chromosome X- and 1-enriched sequences from this study are summarised in table 5. Examples of the BLAST output from dog sequences showing significant similarity to database entries are given in appendix 3.

Table 5 Summary of BLAST database searching for clones from chromosome enriched libraries

Key: PHKA1 = phosphorylase kinase, alpha 1 gene; UTX = ubiquitously transcribed tetratricopeptide repeat gene on chromosomes X and Y; DMD = dystrophin; ABC7 = ATP-binding cassette 7-iron transporter gene; SINE = short interspersed nucleotide element; LINE = long interspersed nucleotide element; MER = medium reiteration sequence; LTR MaLR = long terminal repeat mammalian retrotransposon-like superfamily.

	Chromosome X	Chromosome 1
Clones with putative gene matches	4 x24 - PHKA1 x57 - UTX x74 - DMD x92 - ABC7	0
Clones matching repetitive elements		
SINEs	12	10
LINEs	5	4
Others (MER, LTR MaLR etc.)	4	1
Total	21	15
Clones with matches to known chromosome X-derived material from other species	9	n/a
Clones with no significant matches	52	40

The matches of sequences to SINEs were all to database entries containing a dog (or probably more accurately, a carnivore) SINE, since they all were to dog, seal, and/or mink sequences. The LINE matches were typically to many species, including man, cat, pig, and cattle indicating that this repeat motif is not restricted to any order or species. The other repeat motifs, MER, LTR MaLR, were from human sequences in the GenBank database, which may reflect the content of the database rather than have any biological significance.

3.8 Fluorescence *in situ* hybridisation (FISH)

A minimum of 30 chromosomes were examined for each of the microsatellite-containing clones. At least ten representative digital images were captured and processed with a filter that revealed enhanced DAPI bands. The DAPI counterstain appears blue under UV light, with the detection system used for biotin (Texas Red) producing red signal and that for digoxigenin (FITC) producing green signal.

The small size of the probes gave rise to more background and non-specific signal than experienced with the larger insert dog cosmid and BAC libraries (M. Breen personal communication), although two clones of less than 1kb did give reasonably clean and reproducible signal (see section 2.8.1). The probes used had been selected to contain microsatellites and had been shown by database searches to contain other repetitive elements. Sonicated dog DNA was used as an unlabelled competitor to reduce the non-specific binding of the probes to such repetitive elements. However, even after pre-hybridisation with the competitor DNA, many probes produced speckling caused by non-specific binding, presumably because of the small probe size combined with the relatively high repetitive content (Trask, in Birren *et al.* 1999, p314).

In the majority of experiments, one chromosome 1 probe and one chromosome X probe were hybridised simultaneously. In these cases, the red and green signals were separated digitally and analysed independently.

3.8.1 Chromosome X

Forty-six microsatellite-containing plasmid clones from the X-enriched library were analysed by FISH. Of these, 37 could be unambiguously positioned on chromosome X. In addition, one clone (1-19) from the chromosome 1-enriched library was found to hybridise to chromosome X. Typical images are shown in figures 11 to 17. Chromosome X is clearly identifiable, as it is the only metacentric dog chromosome.

Figure 11 shows a female metaphase spread probed with clone x57. Clear green signal is present on both chromatids of each X chromosome. The single spots on the autosomes represent non-specific hybridisation (speckling); such signals were not reproducible between spreads.

An example of an X-specific probe (x71) hybridised to a male spread is given in figure 12. This clone had given a faint secondary band on digestion (section 2.6.2), but gave no consistent signal on any chromosome other than X.

Figure 13 shows a probe (x51) that hybridised consistently to chromosome X. There are also some signals, typically near the centromere, on single homologues of several autosomes. This pattern of autosomal hybridisation was observed several times which may suggest that the clone, although X-derived, contains a repetitive element common to other chromosomes.

Figures 14a - h shows a series of isolated X-chromosomes or X-Y pairs (from male spreads). Figures 14a and b (clones x21 and x13) show probes that are pseudoautosomal, hybridising to Xp24 as well as the Y chromosome. The remainder (figures 14c to h) hybridise to various positions along the X-chromosomes. Of these, two (x92 figure 14f and x24 figure 14g) showed sequence homology to two genes (ABC7 and PHKA1 respectively). This allows these genes to be tentatively positioned on the dog X-chromosome for the first time. Signal intensity is not consistently related to the size of the clone: x68 (figure 14d) is only 686bp, whereas x24 (figure 14g) is approximately 6kb. Instead, signal intensity presumably relates to target size: x68 is assumed to hybridise to some locally repeated sequence. Signal intensity also depends on chromosome morphology - the assignment of a probe to a band requires at least 30 chromosomes to be examined to allow for the variation in signal, chromosome elongation and variation in the quality of the metaphase spreads. Chromosomes are three-dimensional and the preparation and fixing of them to slides can cause distortions, reducing or eliminating signal from one or both chromatids. Figure 14h shows a clone (1-19 - a clone from the chromosome 1 library) where one chromatid gives a weaker signal than the other, possibly due to the position or bending of the chromosome.

Clone x28 (figure15) hybridised to chromosome X, but often gave signal on only one of the X-chromosomes of a female spread. The reason for this is unclear. Restriction digestion of the clone (data not shown) had produced one insert band of approximately 8kb. The observed location of the clone on chromosome X, suggested that it could lie in the pseudoautosomal region. The linkage data (below) indicated that it was pseudoautosomal or autosomal in origin since heterozygous alleles were present in male animals. If the clone was chimaeric, the contaminating fragment may not have been easily detected by FISH analysis, particularly if it was a smaller fragment than the one seen on chromosome X and had a small local target on the autosome. Likewise, because of the reduced intensity of staining of small fragments on

agarose gels, it may have not have been visible on the gel of the restriction digest of the clone. The strong signal visible in the interphase nucleus in this figure was typical of FISH results from this library; it is presumably due to less effective blocking of repetitive sequences in the dense nucleus.

Figure 11 FISH of clone x57 to female metaphase spread.

Yellow arrows indicate signals on both chromatids at Xp11 – 12 of each X chromosome.

Figure 12 FISH of clone x71 to male metaphase spread.

Yellow arrows indicate signals on both chromatids at Xp21 of the X chromosome.

(Y chromosome is indicated with a pink arrow)

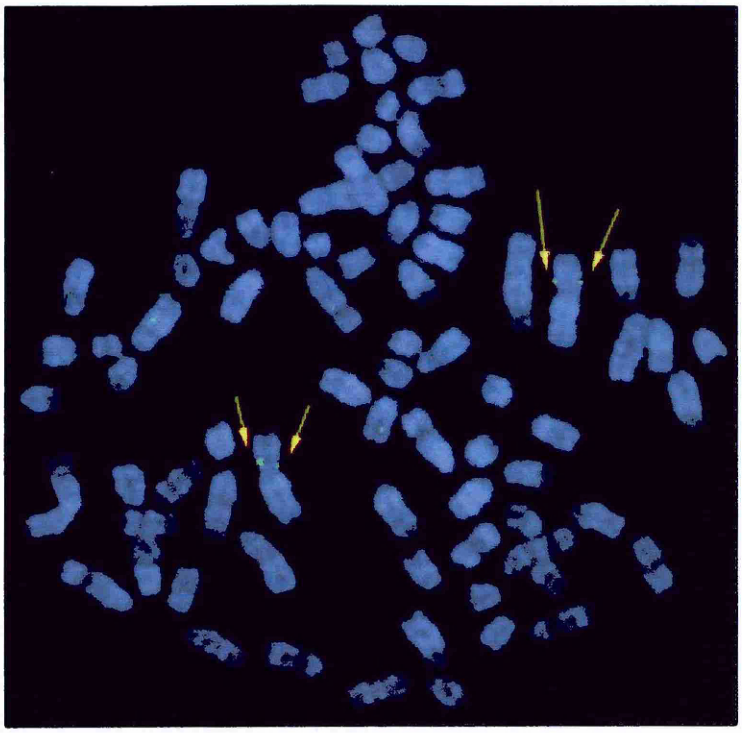


Figure 11

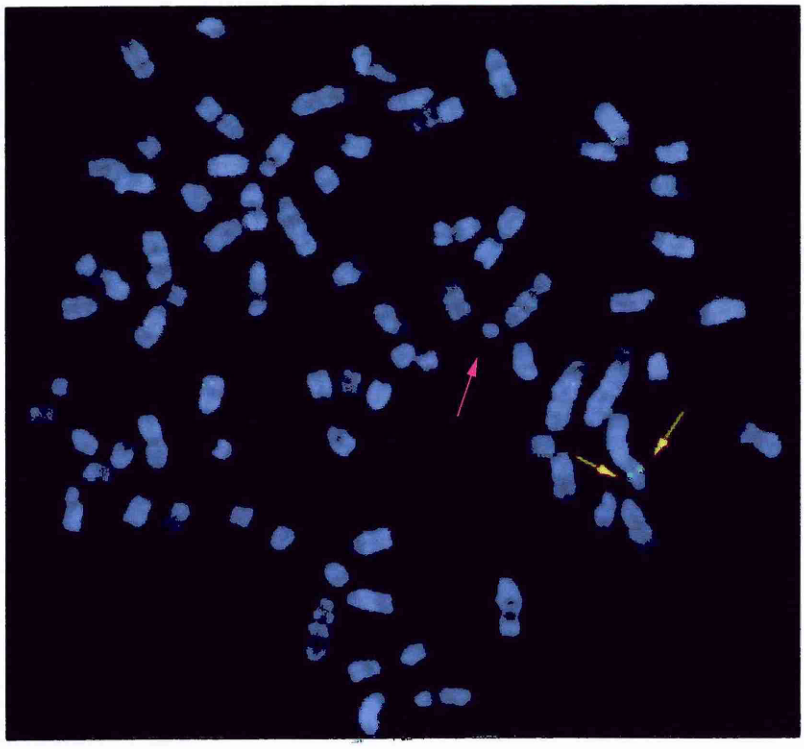


Figure 12

Figure 13 FISH of clone x51 to female metaphase spread.

Yellow arrows indicate signals on both chromatids at Xq22 – 23 of each X chromosome.

Figure 14 Examples of FISH of eight clones showing isolated X or X/Y pairs of chromosomes.

Figure 14a Probe x21 (male metaphase; signal in Xp24, Y chromosome is shown on the right)

Figure 14b Probe x13 (male metaphase; signal in Xp24, Y chromosome is shown on the right)

Figure 14c Probe x78 (female metaphase; signal in Xp 23 – 24)

Figure 14d Probe x68 (female metaphase; signal in Xq26 – 27)

Figure 14e Probe x27 (female metaphase; signal in Xq13 – 21)

Figure 14f Probe x92. (female metaphase; signal in Xq12)

Figure 14g Probe x24 (female metaphase; signal in Xq12)

Figure 14h Probe 1-19 (female metaphase; signal in Xq11 – 12. This clone was derived from the chromosome 1 library but was found, by FISH mapping to be located on chromosome X; no signal was observed for this probe on any other chromosome).

Figure 15 FISH of clone x28 to female metaphase spread.

Yellow arrows indicate signals at Xp23 - 24 on both chromatids of one X chromosome. Pink arrow indicates the other X chromosome. An interphase nucleus is visible on the right.

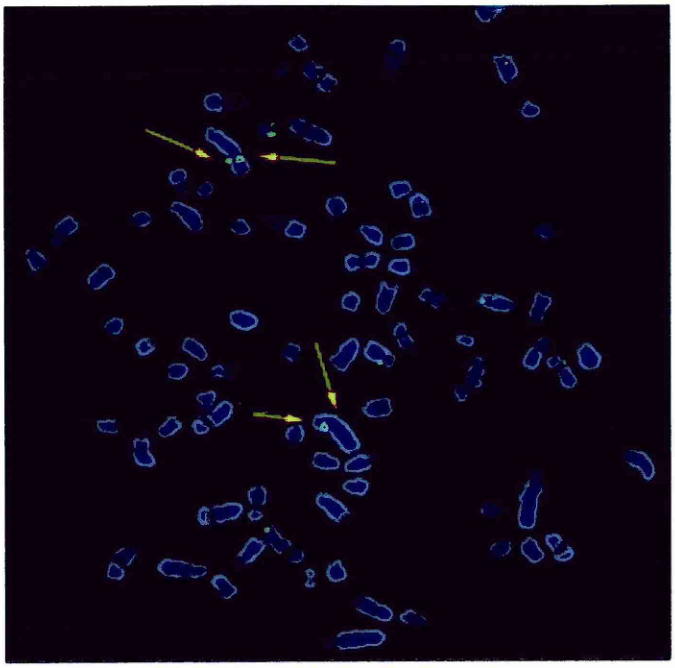


Figure 13

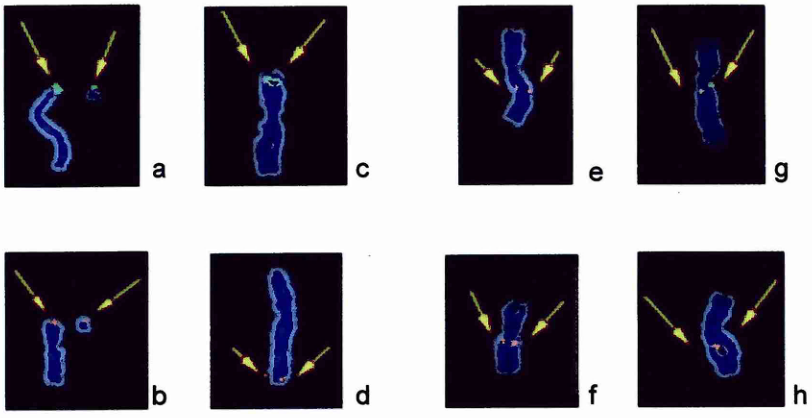


Figure 14

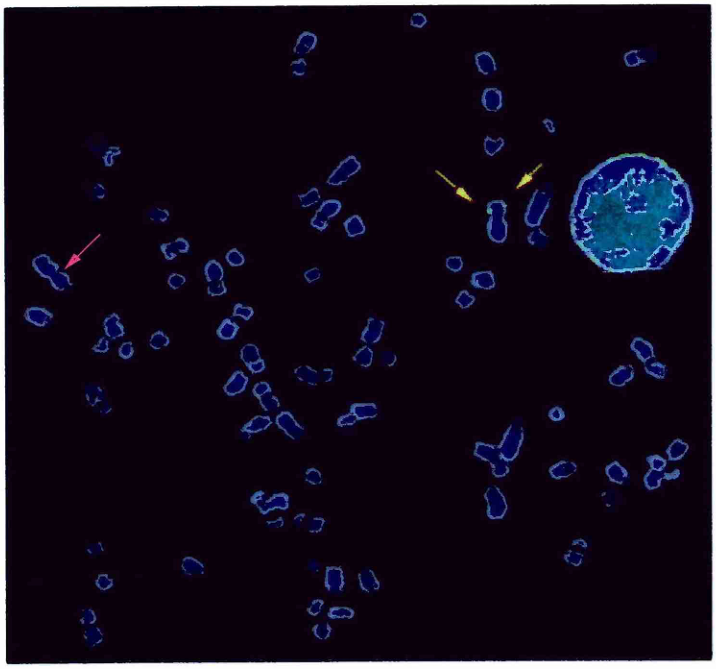


Figure 15

Figure 16 Dual colour FISH of two probes in the DMD region.

The metaphase spread is from a female carrier of DMD. Signals from x74 (red) and x71 (green) are visible on both chromatids of one X chromosome. No signal is visible on the other X chromosome (pink arrow), which is presumed to carry a deletion in this region.

Figure 17 Dual colour FISH two probes to a male metaphase spread.

Signal from probe x49 (green) at Xq11, signal from x77 (red) at Xq13; indicated by yellow arrows (Y chromosome is indicated with a pink arrow)

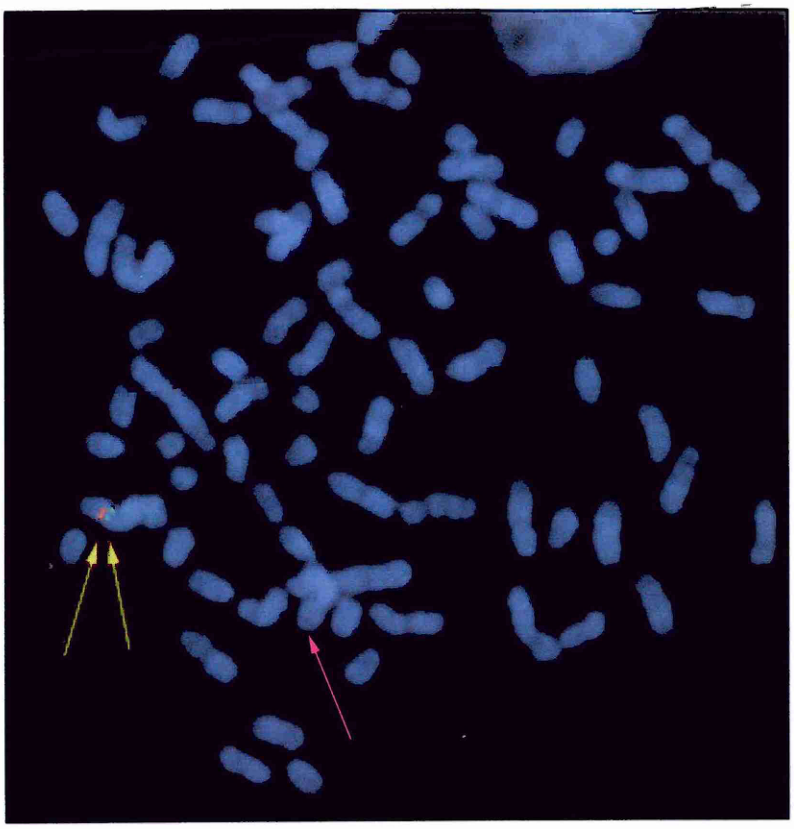


Figure16

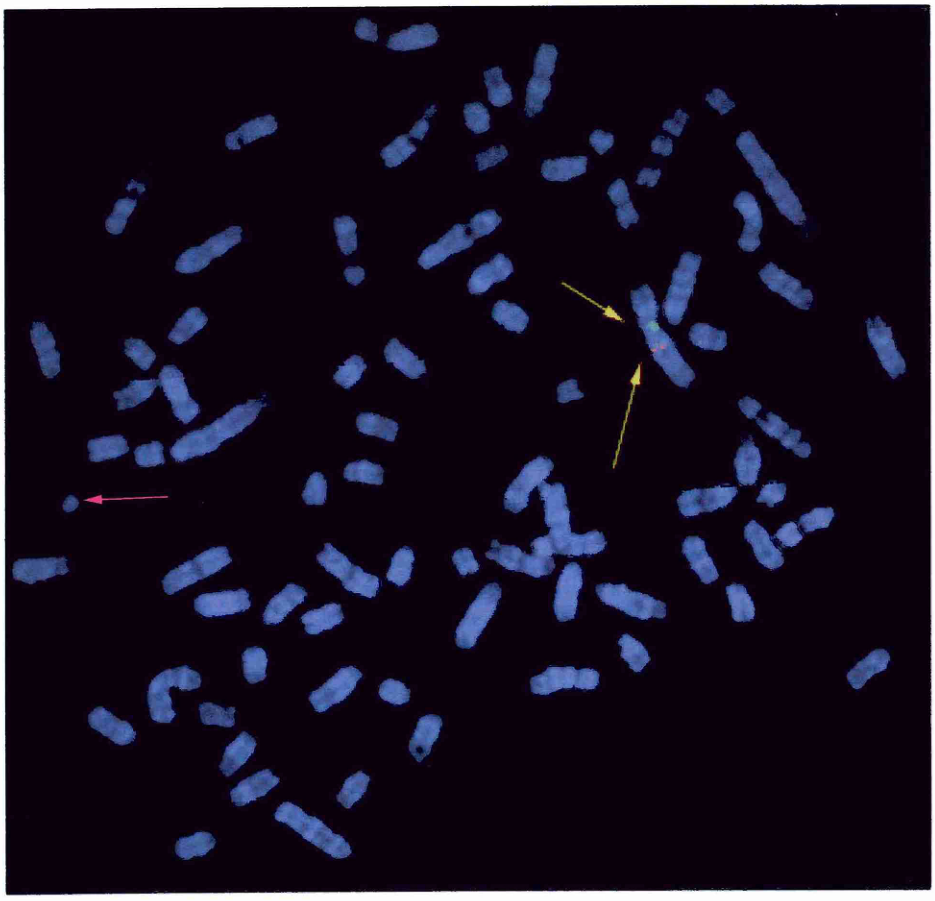


Figure 17

In some FISH experiments, two probes from the same chromosome were co-hybridised but analysed separately. Figure 16 shows the dual colour detection of two probes, x74 and x71. The metaphase spread was derived from a female dog known, from breeding experiments, to be heterozygous for a deletion in Xp21, a region containing the dystrophin locus. The spread shows the normal X chromosome with both probes producing signal, while the other X chromosome has no signal. Database analysis of the sequence of clone x74 indicated homology with part of the gene sequence for human Duchenne muscular dystrophy. The other probe shown (x71) has no apparent database match and represents an STS located in the region shown to be deleted in the muscular dystrophy carrier animal. (In addition, it can also be seen that this probe only hybridises to the X chromosome, although it was digested to two bands with *Hind*III.) This work was part of a larger collaborative project, the results of which have been published (Schatzberg *et al.* 1999) and a reprint of the paper is included at the end of this thesis.

Figure 17 shows dual colour detection of two probes, x49 and x77 on a male metaphase spread. Signals from both probes are visible on both sister chromatids of the X chromosome. This is an example of a clean metaphase preparation with no cellular debris and well-banded chromosomes, making identification of the location of these clones straightforward.

A further 21 probes gave reproducible and specific hybridisation to various regions of the X-chromosome (FISH data not shown, cytogenetic locations given in table 6).

Clones x12, x20, x53, x55 and x94 produced some signal on the X-chromosome, but not reproducibly. Therefore, these clones could not be definitively mapped to chromosome X by FISH. They may contain repetitive elements, which are located elsewhere in the genome as well as on chromosome X, and therefore the signals produced were too diffuse to define a region of hybridisation on either chromosome X or the autosomes.

A summary of the cytogenetic location of the 36 chromosome X-derived probes that were unambiguously mapped is given in table 6 and illustrated in figure 18. Figure 18 also shows the position of 1-19, a clone derived from the chromosome 1 library but found, by FISH analysis, to be located on chromosome X.

Table 6 Cytogenetic position of chromosome X library clones

Cytogenetic locations, as determined from 30 metaphase spreads in each case, are given. PAR = pseudoautosomal; region; PHKA1 = phosphorylase kinase alpha 1; UTX/UTY = ubiquitously transcribed tetratricopeptide repeat gene on chromosome X / Y; ABC7 = ATP-binding cassette 7-iron transporter.

Clone name	Cytogenetic location	Comments
x1	Xq24 - 25	
x4	Xp13 - 21	
x7	Xp21 - 22	
x9	Xp13	
x13	Xp24 & Y	PAR
x21	Xp24 & Y	PAR
x24	Xq12	Similarity to PHKA1 from rabbit, human & mouse
x27	Xq13 - 21	
x28	Xp23 - 24	
x30	Xp12 - 13	
x31	Xp23	
x33	Xq22 - 23	
x34	Xp11 - 12	
x35	Xq23 - 25	
x36	Xp21 - 22	
x38	Xq21.3 - 22	
x40	Xq21.3 - 23	
x47	Xp13 -21	
x49	Xq11	
x50	Xq22 -23	
x51	Xp21 - 23	
x56	Xq21.2 - 21.3	
x57	Xp11 - 12	Similarity to UTX & UTY from mouse & human
x58	Xq26 - 28.1	
x64	Xq11	
x66	Xq21.2 - 22	
x67	Xp13 - 21	
x68	Xq26 - 27	
x69	Xq21.1 – 21.3	
x71	Xp21	Deleted in DMD dogs
x73	Xp11 - 13	
x74	Xp21	Homology with human dystrophin, deleted in DMD dogs
x77	Xq13	
x78	Xp23 - 24	
x91	Xp22 - 23	
x92	Xq12	Similarity to ABC7 from mouse & human

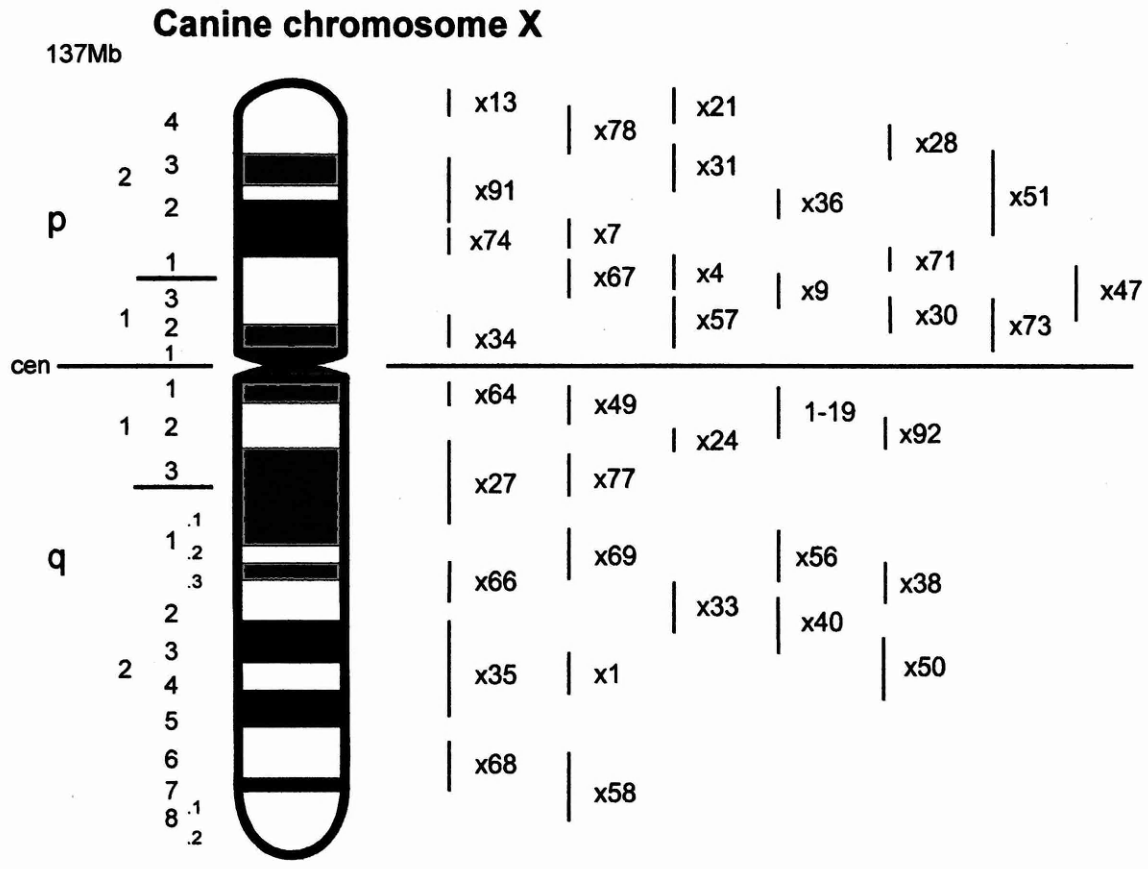


Figure 18 Results of FISH of chromosome X-enriched clones

Chromosome X in the dog is described as a sub-metacentric chromosome, with the smaller p arm being approximately a third of the length of the chromosome and the longer q arm approximately two thirds of the length. The ideogram shown is based upon that of Reimann et al. (1996). The size of chromosome X is taken from Langford et al. (1996). Clone names are marked alongside bars indicating the range over which signal was seen and include marker 1-19 derived from the chromosome 1-enriched library but found to be located on chromosome X.

3.8.2 Chromosome 1

Figures 19 to 23 show examples of hybridisation of microsatellite-containing clones (putative and confirmed) from the chromosome 1-enriched small insert library. The FISH analysis of 47 chromosome 1 library-derived clones enabled 23 to be mapped.

In several of these instances, signal is seen on only one chromatid (e.g. figure 19). Repeated observations, however, more often show signal on both chromatids. The absence of signal on one chromatid is presumed to be due to the morphology of the chromosomes, and exemplifies the need to examine a representative number of spreads.

Clone 1-33 (figure 20) is smaller (approx. 0.6kb) than would be expected to give a strong signal. As with x68 (section 2.8.2), this may be due to close repetition of the target sequence creating a larger potential region for hybridisation. However, database searches with end-sequence from this clone revealed no known repetitive elements.

Figure 21 shows clone 1-18 hybridising in 1q13.2 – 13.3 with no consistent signal observable elsewhere, despite the fact that digestion with *Hin*dIII produced two fragments of approximately 8kb and 5kb (section 2.6.2). This implies that either one *Hin*dIII fragment is giving no signal, or that the clone is a result of incomplete digestion of the flow-sorted chromosomes, rather than co-ligation of two unrelated fragments, or that the largest band generated during restriction digestion was a result of partial digestion of the plasmid (hence the observation of signal in only one location). The two signals given by the probe on one of the chromosomes in the spread used in figure 21 are skewed - this is presumed to be an effect of chromosome morphology in fixed preparations, and is not uncommon.

Individual examples of chromosomes hybridised with probes from the chromosome 1-enriched library are shown in figure 22 a to f . In all these examples, signal is present on both chromatids and was observed on both chromosomes in a representative sample for each probe.

Figure 23 shows a complete metaphase spread with clone 1-46 hybridising to chromosome to 1q12, showing two signals on one chromosome but on only one chromatid of the other.

Table 7 gives a summary of the cytogenetic location of the 23 probes that were unambiguously mapped and these are illustrated in figure 24.

Figure 19 FISH of clone 1-32 to metaphase spread.

Yellow arrows indicate signals at 1q34 – 35.2 on both chromatids of one homologue of chromosome 1, and on one chromatid of the other.

Figure 20 FISH of clone 1-33 to metaphase spread.

Yellow arrows indicate signals at 1q31.2 – 31.3 on both chromatids of one homologue of chromosome 1, and on one chromatid of the other. Part of an interphase nucleus is visible at the top of the image.

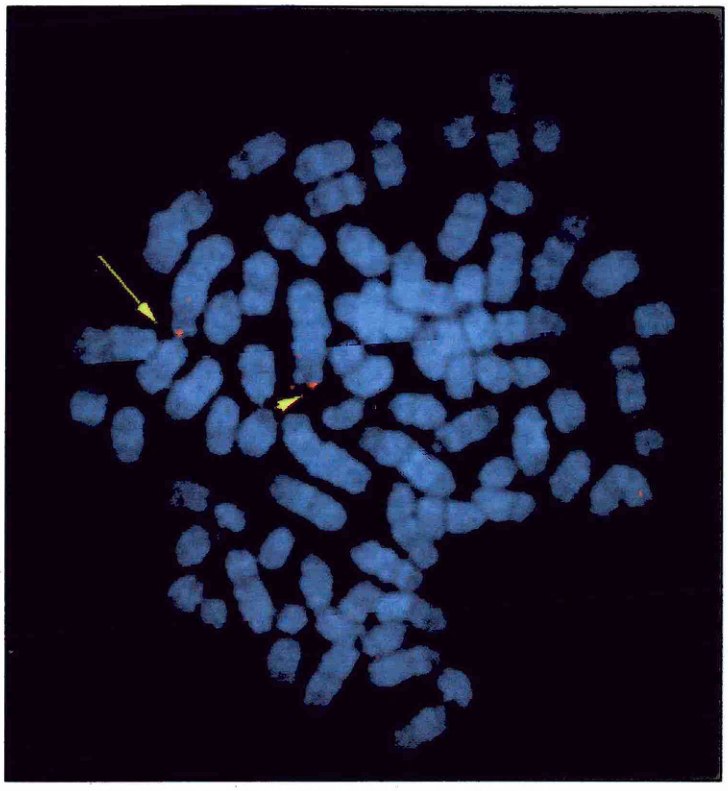


Figure 19

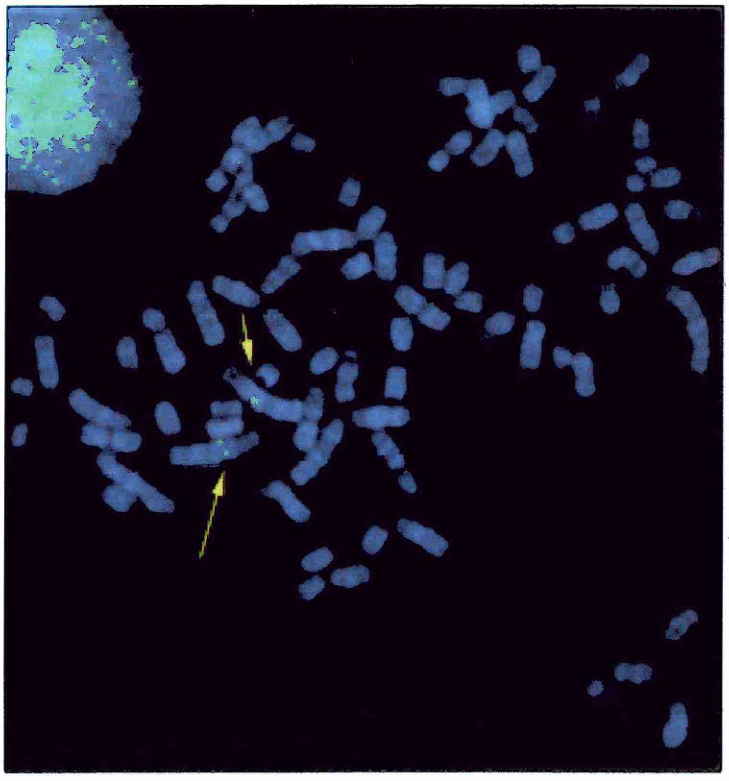


Figure 20

Figure 21 FISH of clone 1-18 to metaphase spread.

Yellow arrows indicate signals at 1q13.2 – 13.3 on both chromatids of one homologue of chromosome 1, and on one chromatid of the other.

Figure 22 Examples of FISH of six clones showing isolated chromosomes.

Figure 22a Probe 1-34 (signal in 1q31.1 – 31.3)

Figure 22b Probe 1-60 (signal in 1q12 – 13.1)

Figure 22c Probe 1-28 (signal in 1q31.1 – 31.2)

Figure 22d Probe 1-43 (signal in 1q13.3 – 14.1)

Figure 22e Probe 1-01 (signal in 1q13.3 – 14.1)

Figure 22f Probe 1-57 (signal in 1q31.1 – 31.2)

Figure 23 FISH of clone 1-46 to metaphase spread.

Yellow arrows indicate signals at 1q12 on both chromatids of one homologue of chromosome 1, and on one chromatid of the other.

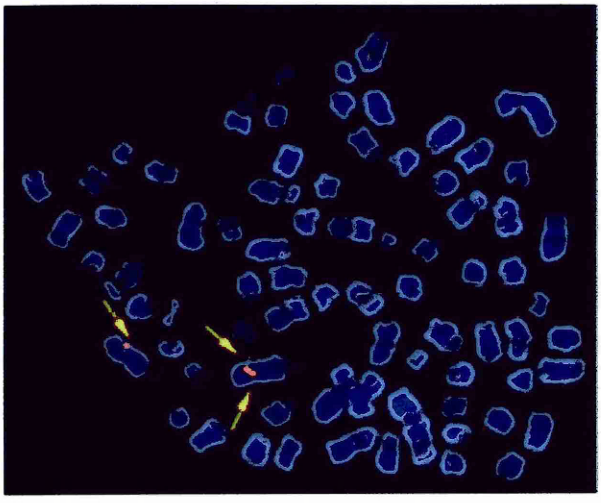


Figure 21

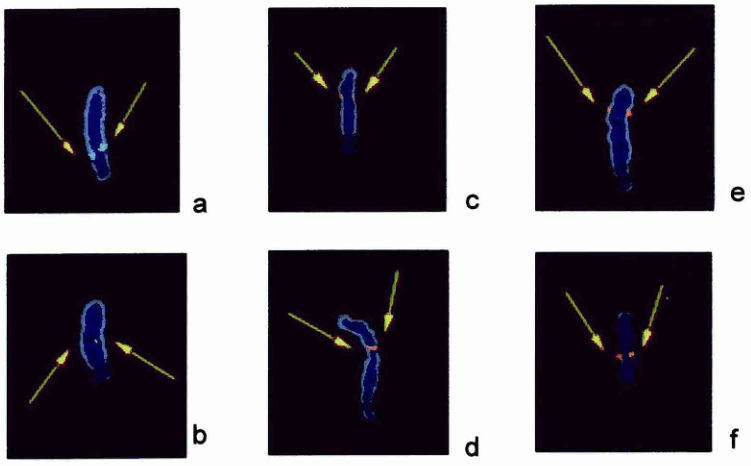


Figure 22

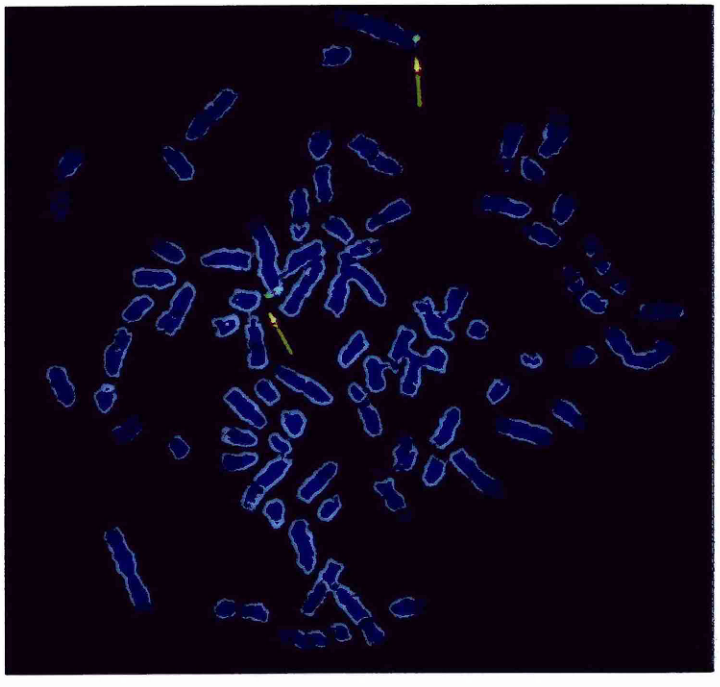


Figure 23

Table 7 Cytogenetic position of chromosome 1 library clones

Clones from the small insert chromosome 1 library were hybridised to the cytogenetic locations shown below. No significant database matches were found for the sequences from any of the chromosome 1-enriched library clones, other than to repetitive elements.

Clone name	Cytogenetic position
1.01	1q13.3 - 14.1
1.03	1q14
1.09	1q32
1.11	1q23 - 24
1.18	1q13.2 - 13.3
1.20	1q21 - 23
1.21	1q35
1.27	1q35.2 - 35.3
1.28	1q31.1 - 31.2
1.30	1q14.3 - 21
1.32	1q34 - 35.2
1.33	1q31.2 - 31.3
1.34	1q31.1 -31.3
1.42	1q32 - 33
1.43	1q13.3 - 14.1
1.45	1q33 - 34
1.46	1q12
1.49	1q31
1.51	1q12 - 13.1
1.57	1q31.1 - 31.2
1.58	1q33 - 34
1.59	1q12
1.60	1q12 - 13.1

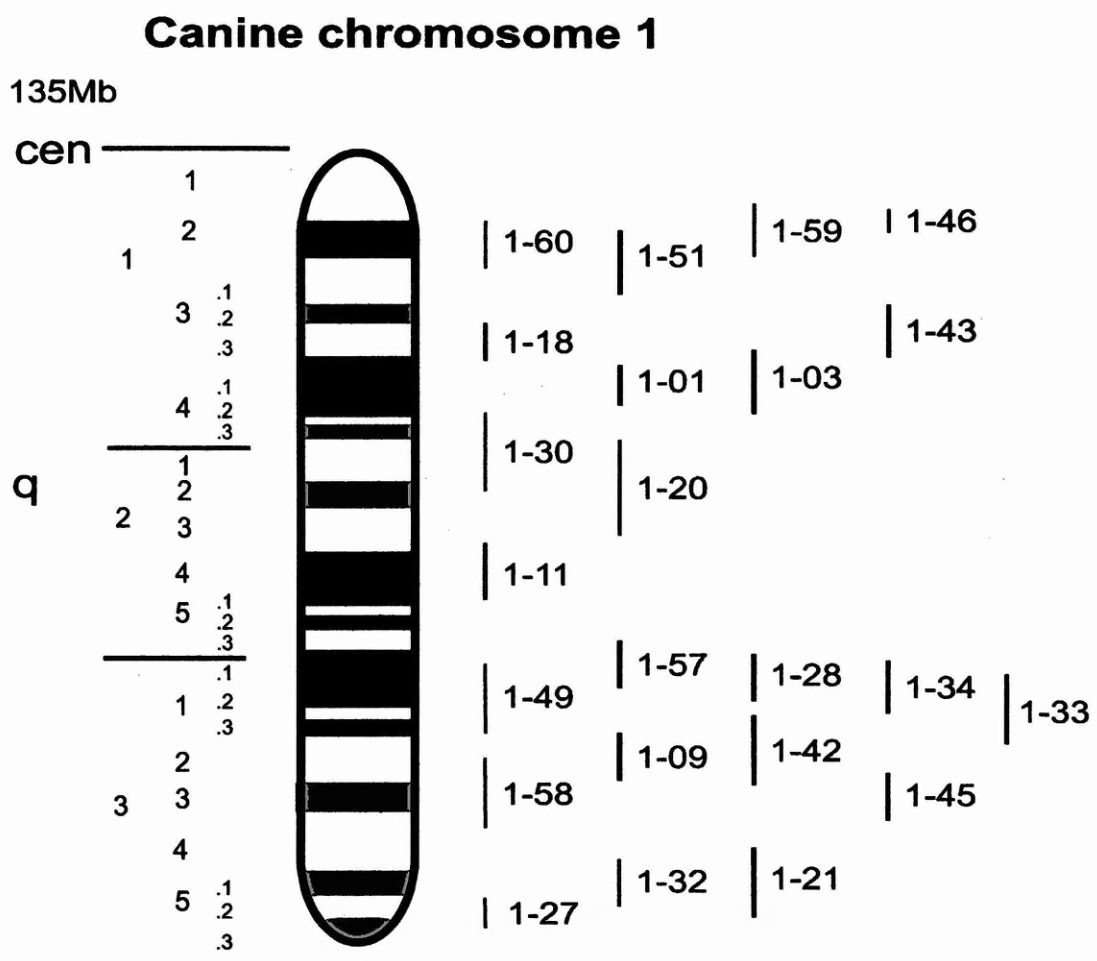


Figure 24 Results of FISH of chromosome 1 -enriched clones

Chromosome 1 in the dog is described as an acrocentric chromosome, with the q arm accounting for the entire length. The ideogram shown is based upon that of Reimann et al. (1996). The size of chromosome 1 is taken from Langford et al. (1996). Clone names are marked alongside bars indicating the range over which signal was seen.

3.8.3 Purity of flow-sorting

The majority of the clones which gave clean, reproducible signals were located on the expected chromosomes, demonstrating that the libraries are significantly enriched for chromosomes X and 1, with respectively 91% and 86% hybridising to the expected chromosome. However, a number of clones were found to hybridise to chromosomes other than X (four) or 1 (eight); these are summarised in table 8.

Clone 1-19, derived from the chromosome 1-enriched library, was assigned by FISH to chromosome Xq11 - 12 and was subsequently used as a chromosome X-specific marker. Clones x54 and x60 hybridised to different autosomes in the range 21 – 38. These smaller autosomes are difficult to unambiguously identify without the use of the dog flow-sorted paint probes. This was not carried out because the paints are a limited resource and the two clones did not appear to be located on autosomes of interest to the dog-mapping project at the AHT. Some contamination of the libraries with material from other chromosomes was expected because flow sorting does not produce pure chromosome isolates. Human flow sorting can achieve purities of up to 95% (Ross & Langford, 1997 In Genome Mapping, Dear, P. H. p172). The purities of the dog chromosome-specific libraries produced in this study are similar to that obtained from flow-sorting human chromosomes.

Assignment of the clones to the contaminating chromosomes is not definitive; therefore, it is not possible to say whether there is any significance to the chromosomes that co-sorted with X and 1. Flow-sorting errors are likeliest between similar chromosomes, such as X and 1. It is possible that combinations of fragments or of smaller chromosomes combined to give signals similar to chromosomes X or 1, and were therefore co-sorted.

Table 8 Contaminant clones in flow-sorted, small insert libraries

The clones that consistently hybridised to chromosomes other than those expected are shown with their putative locations.

Clone name	Cytogenetic location
x15	2q12.1 – 12.3
1-39	2q31 – 33
1-08	4q23 –24.1
1-37	4q23
x18	7q?
1-13	10q12 – 14
1-48	10q?
1-47	12q
1-31	18q21
x54	21 - 38
x60	21 - 38
1-19	Xq11 - 12

3.9 Linkage analysis

The chromosome 1 markers were not included in this part of the study because there was insufficient time when the reference family was released to examine a large number of markers. Furthermore, it is likely that at least one of the existing linkage groups in the published maps would be located on chromosome 1 and on-going work at the AHT was addressing that issue concurrently with this study. The chromosome X markers were given priority since they represented a large resource of physically assigned markers on a chromosome where only four markers had previously been published.

Primers were designed to flank microsatellite markers from the chromosome X-enriched library (as described in 2.9) to produce an outline genetic linkage map of the chromosome. A cosmid clone (H62) known to contain a microsatellite, to which primers had already been designed, was discovered to map to chromosome X by co-workers at the Animal Health Trust (N. Holmes, personal communication), and was included in this study.

Microsatellite markers associated with the X-linked genes F8c, F9, PGK1, MNK and CHM were also tested for suitability for mapping. These primers were obtained from published work or provided by colleagues involved in dog genome mapping:

F8c and F9 – E. Ostrander, personal communication,

PGK1 and CHM – Deschênes *et al.* (1994).

MNK – Venta *et al.* (1996).

The DogMap animals were part of the original mapping effort but were limited in only being two-generation, phase-unknown pedigrees and were therefore not optimal for linkage analysis. The initial linkage map (Lingaas *et al.* 1997) was produced with these animals and, at the beginning of this study, they were the only available mapping resource. Preliminary work was therefore carried out on these animals. The Cornell reference families were made available at the end of 1998 (the pedigree structure is shown in figure 3). They are the most powerful mapping tool currently available and benefit from being phase-known, although they are actually interlinked, individuals from each family having been used to breed with members of the other families. The published linkage maps of Mellersh *et al.* (1997) and Neff *et al.* (1999), have used a much larger extended pedigree from the Cornell and other families. Neff *et al.* (1999), used 218 animals from the Cornell families, to type 276 markers including the markers in the first map

published using these reference families (Mellersh *et al.* 1997). Five of the markers examined were X-linked: C0X.314, AR, PGK1, CHM and F8c.

Thirty-five markers from the chromosome X library, five X-linked gene markers and one cosmid clone located on X (H62, M. Breen, personal communication) were examined to establish which were polymorphic. Initially, the markers were typed on the DogMap parents and other, unrelated animals, to establish polymorphism. However, when the Cornell families were released, they were typed on the parents from these families. This showed that 22 markers were polymorphic in the Cornell families, although five of these were only polymorphic between families and so were uninformative. The remaining 19 were monomorphic, although results from the preliminary study with DogMap and other animals indicated that five of these, though monomorphic in the Cornell animals, were polymorphic in a wider population. Table 9 summarises the results of testing the markers for polymorphism in the Cornell families.

Table 9 Polymorphism of markers in the Cornell families.

Markers that were monomorphic in the Cornell families, but polymorphic in other families, are indicated by an asterisk.

Polymorphic & informative	x9 x12 x20 x21 x24 x27 x28 x50 x51 x58 x64 x66 x68 H62 F8c F9 PGK1
Polymorphic but not informative	x1 x33 x53 x67 x91
Monomorphic	x26 x30 x31 X34 x35* x40 x48 x49 x55* x56* x57 x69 x71* x73 x74 x77* x78 CHM MNK

The 17 markers informative in the Cornell parents were typed, as described in section 2.11, on the relevant families. Of these, five showed autosomal or pseudoautosomal inheritance (from the observation of heterozygous alleles in male animals), whilst the remainder showed typical sex-linked segregation; this is summarised in table 10. Examples of typical genotyping traces are given in figures 25a, b and c. Figure 25a shows some of the data for marker x12.

These primers amplified two alleles in the Cornell families of 137 and 142bp. Male dogs amplified two alleles indicating that this marker is located in the pseudoautosomal region or is from an autosome. FISH analysis for this clone did not provide an unambiguous location on

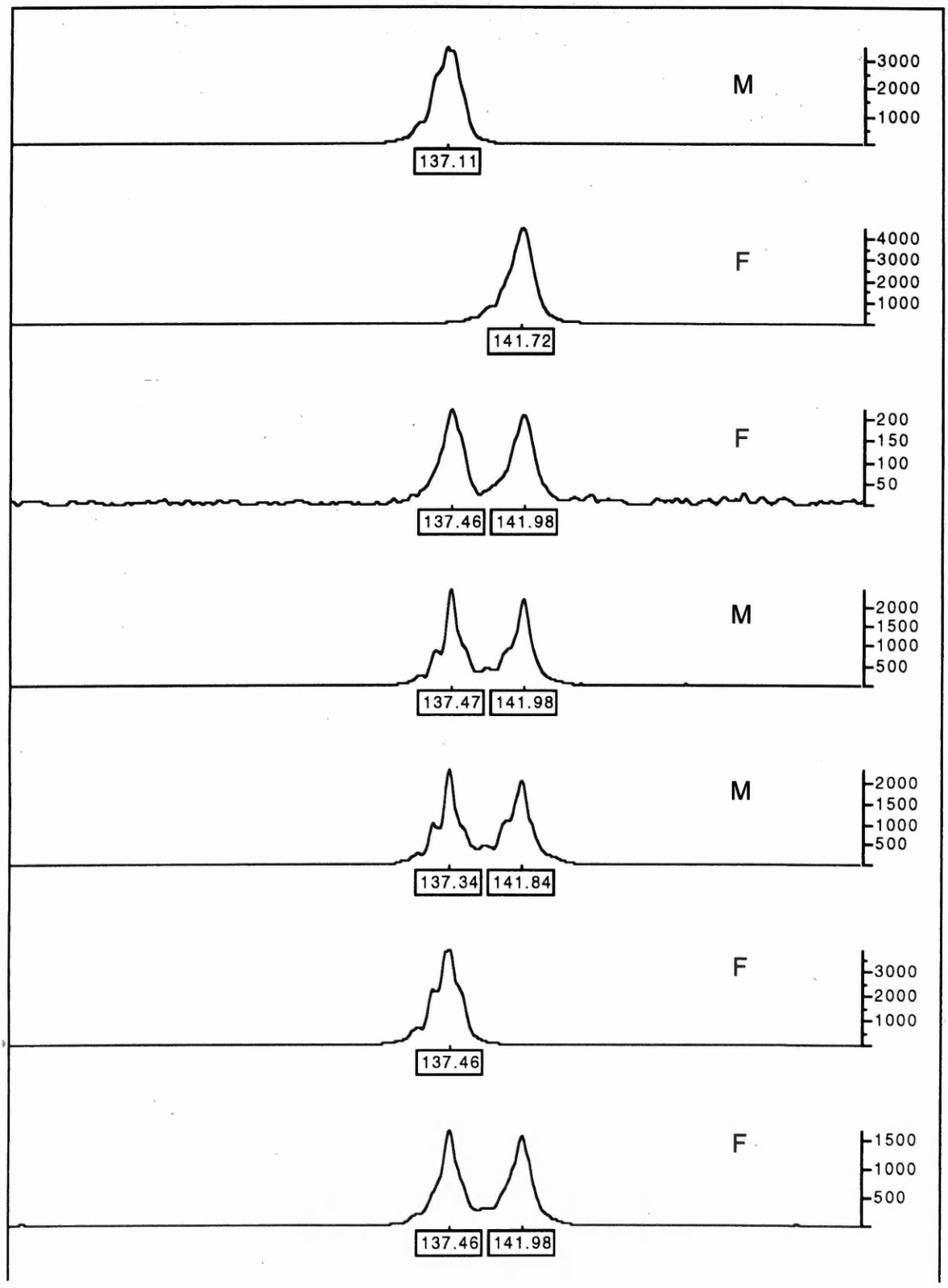


Figure 25a Genotyping trace from x12 $(CT_4)GT(CT_7)$ with members of the Cornell families. *This primer amplified two alleles in male animals, FISH did not provide an unambiguous location for the clone, and the marker did not show linkage to any other marker. This marker showed two alleles in the Cornell families of 137 and 142bp. F = female, M = male*

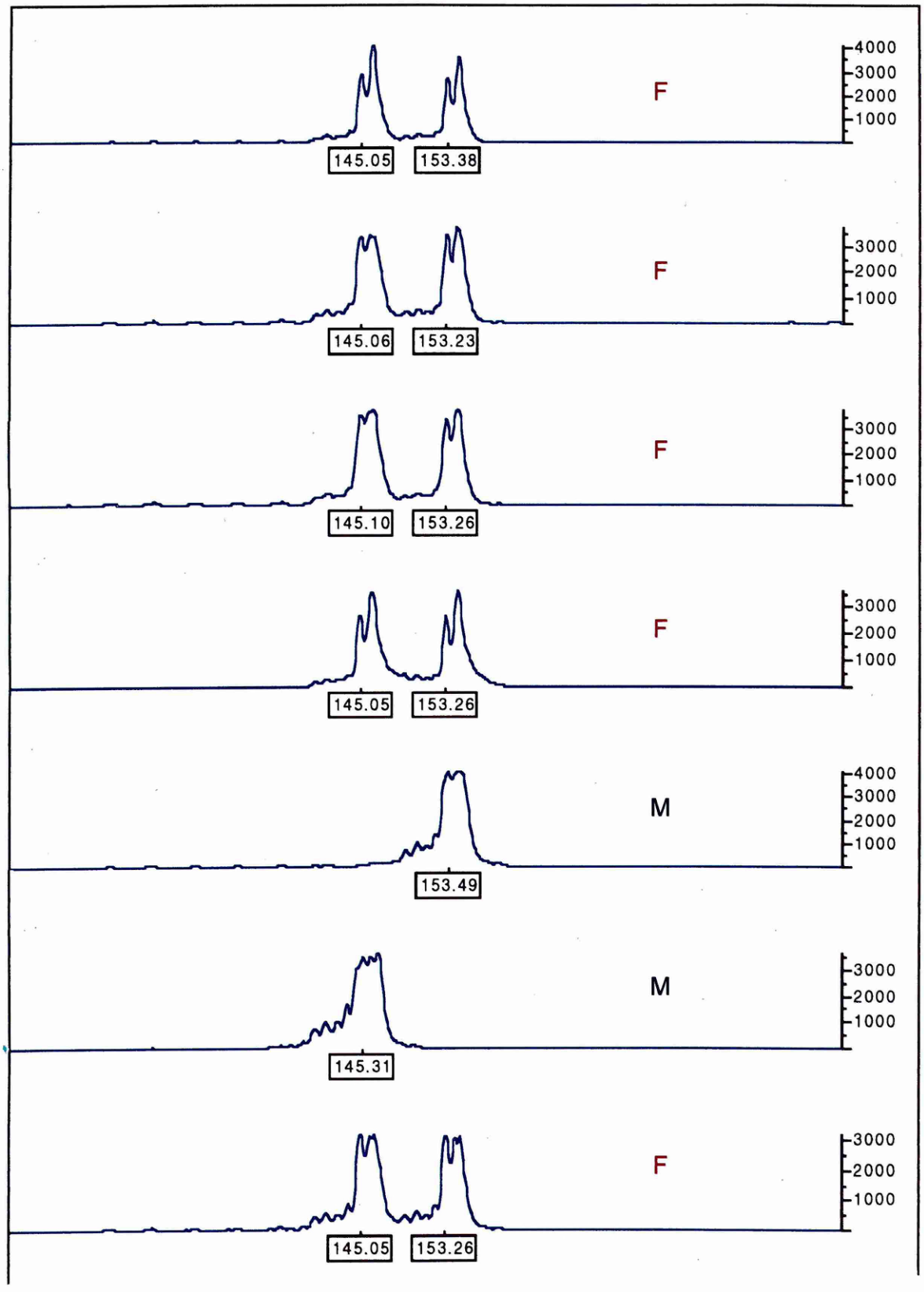


Figure 25b Genotyping trace from F9 $(TTTA)_n$ with members of the Cornell families. *This marker showed two alleles in the Cornell families of 145 and 153bp. It was linked to 3 other X-linked markers H62, x50 and x58. F = female, M = male*

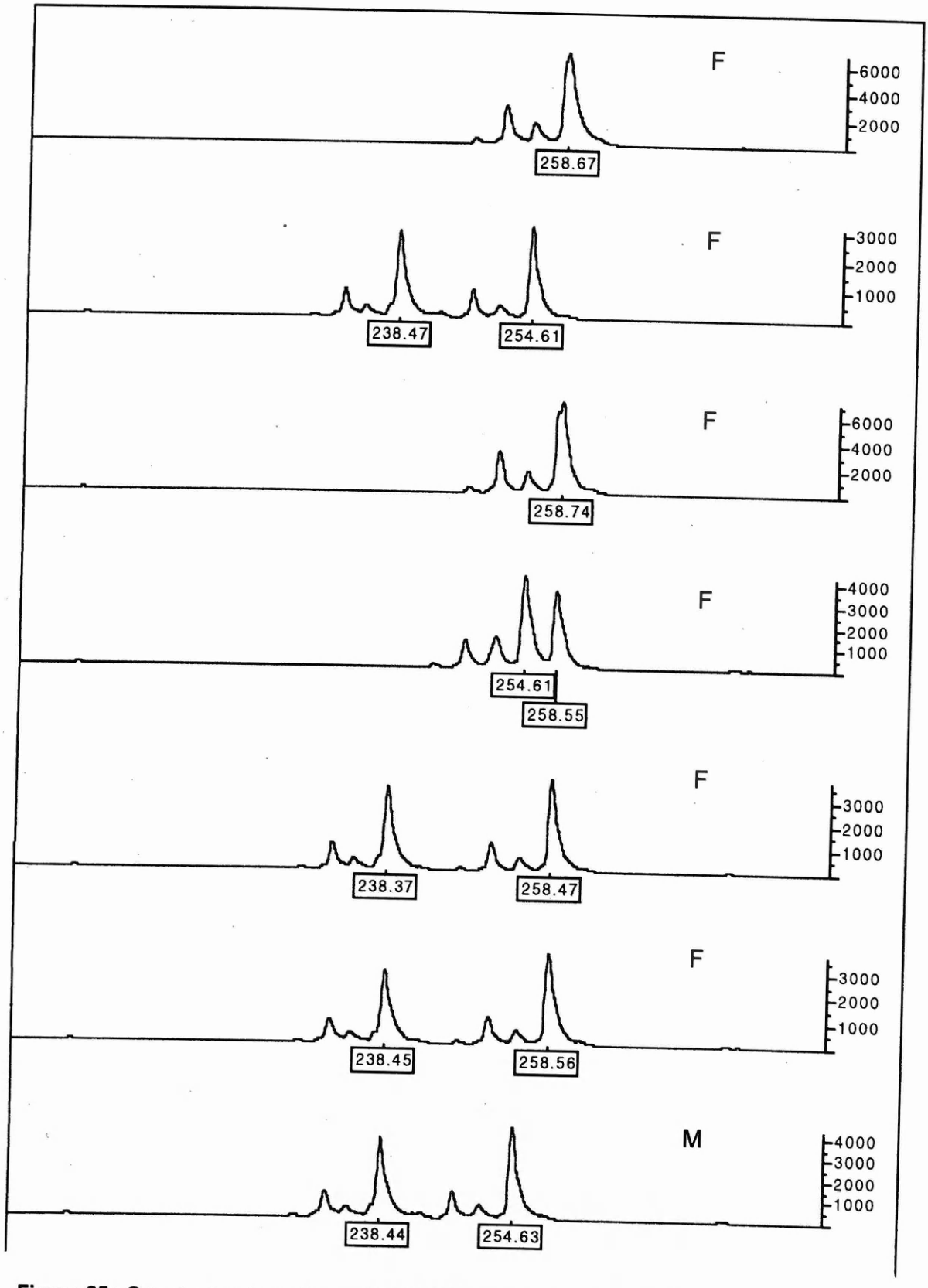


Figure 25c Genotyping trace from x28, $(CTTT)_{20}$ with members of the Cornell families. *This primer amplified two alleles in male animals, FISH located the clone in Xp24 and the marker showed linkage to markers in linkage group 7 of Neff et al. (1999). This marker had 7 alleles in the Cornell families, from 230 to 258bp F = female, M = male*

chromosome X or any of the autosomes. Figure 25b shows data for F9 (clotting factor 9). This marker amplified 2 alleles of 145 and 153bp in the Cornell animals and was inherited in a typically sex-linked manner. Linkage analysis indicated the marker was linked to three other, sex-linked markers, H62, x50 and x58. Figure 25c shows results from marker x28. This marker also appeared to be either pseudoautosomal or autosomal since two alleles were amplified in some of the male members of the families. FISH analysis indicated that this marker was located in Xp24. This marker was very polymorphic, with 7 alleles from 230 to 258bp amplified in the Cornell families.

The genotyping data from x12, x20, x21, x28 and x51 indicated that they were pseudoautosomal or autosomal in origin since heterozygous alleles were present in male animals; or that the primers amplified a repetitive element present in more than one location which was not apparent during initial sequence analysis. Of these, x12 and x20 did not give consistent FISH results to chromosome X or any other autosome and were not able to be physically assigned, nor did they show linkage to any other markers. It is likely that these clones are located on autosomes and further FISH experiments would be able to determine whether this is the case or not. The three other markers, x28, x21 and x51 had shown reproducible signal on the X chromosome, in Xp23 – 24. One of these, x21, had been hybridised to male metaphase spreads and shown signal at the telomere of the Y chromosome as well as in Xp24. Data from both meiotic linkage mapping and FISH therefore suggest that x21 was within the pseudoautosomal region. The remaining 12 markers produced results consistent with being sex-linked but were frequently informative in only a subset of the animals in the families. This limited the data analysis and resulted in few markers showing linkage because they shared few meioses in common, as demonstrated by table 10.

Table 10 Informativeness of markers with respect to Cornell family mapping

Of the 17 markers that were polymorphic and informative in the Cornell families, 12 showed sex-linked segregation and five were heterozygous in male animals indicating pseudoautosomal or autosomal segregation. The families that were informative for each marker, and the number of heterozygotes, are shown.

Marker	X-linked	Informative in family: CF4	CF5	CF6	CF7
		No. of heterozygotes			
x27	+	16	14		
x9	+		12	5	8
x58	+	11		6	6
x66	+	9	9		3
F9	+	8		5	7
x68	+	7	13		
F8c	+	7	9		3
H62	+			6	10
x50	+	4			7
PGK1	+				8
x24	+		7		
x64	+	3			
x28	-	20	14	14	19
x12	-	10		13	13
x51	-	18	15		
x21	-			15	
x20	-	4	2		

The genotyping results are given in appendix 4. The initial analysis was performed using the MLINK option of the Linkage programme (Lathrop *et al.* 1984) to generate lod scores from two-point analysis. Markers located on the X chromosome of females can produce homozygous or heterozygous patterns of segregation, whereas in males, sex-linked markers should only produce homozygous patterns. However, if the markers are located in the pseudoautosomal regions of the X chromosome, these regions undergo recombination and can therefore produce heterozygous patterns of segregation in males. Sex-linked markers were analysed separately from markers located in the pseudoautosomal region, which were treated as autosomal markers in the Linkage programme.

Results of MLINK two-point linkage analysis are presented in table 11 for the markers showing significant lods. Lod scores of 2.00 and over are considered significant for sex-linked

markers (Ott, 1991). All other two-point lods were <2.00. (It should be noted that given the limited genotyping data available, a lod of <2.0 does not necessarily indicate that the markers are unlinked: many pairs of markers share few or no meioses, giving little or no opportunity to detect linkage. This is clearly true for markers that segregate in different families, e.g. H62 and x68, but is also true for some other pairs that segregate in different individuals within the same family. (This limitation impedes multipoint analysis; this was attempted but as expected was not successful for this dataset.)

Table 11 Two-point linkage data from chromosome X markers

Lod scores of >2.0 are considered significant between X-linked markers

Markers	Recombination fraction, θ	Lod scores
x58 – x50	0.06	6.93
x58 – x68	0.06	3.19
H62 – x50	0.10	4.79
PGK1 – x50	0.10	3.64
PGK1 – x58	0.10	3.64
F9 – x50	0.11	2.69
F9 – x58	0.11	5.88
PGK1 – x9	0.13	2.20
H62 – F9	0.17	3.57
H62 – x58	0.23	2.36

3.10 Radiation hybrid analysis

The suitability of primers for use with WG-RH mapping was determined by performing PCR with dog genomic DNA, clone DNA (where available) and hamster cell line A23 DNA; an example is illustrated in figure 26.

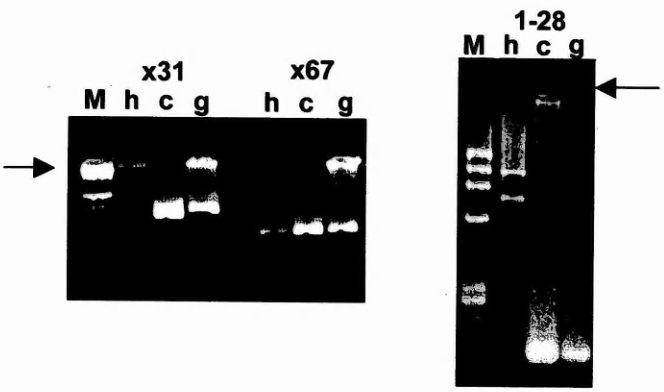


Figure 26 Examples of results of optimisation of PCR products for WG-RH mapping

Key: h = hamster DNA, c = clone DNA, g = dog genomic DNA, M = DNA size standards. Arrows indicate the positions of the sample wells.

Primers from x31 amplified the expected product from the clone and dog genomic DNA only, and hence required no further optimisation. Primers from x67 amplified a PCR product from hamster DNA of a similar size to the dog and clone products; this could not be eliminated by further optimisation. Primers from 1-28 amplified sequences from hamster DNA of a different size to the products from clone and dog DNA; therefore, the co-amplification from hamster DNA did not interfere with WG-RH analysis.

Primers that produced a PCR product of the same size from dog and hamster DNA were subjected to further PCR optimisation including the Sanger PCR protocol, or one or both of the primers were redesigned.

Markers were then typed on the panel, as described in section 2.12. Examples of the results obtained for three markers are shown in figure 27. Two of the markers (1E7 & x21) are linked and a similar, but not identical, retention pattern may be observed from the PCR products; the other marker (x27) is unlinked to the first two and shows a different retention pattern.

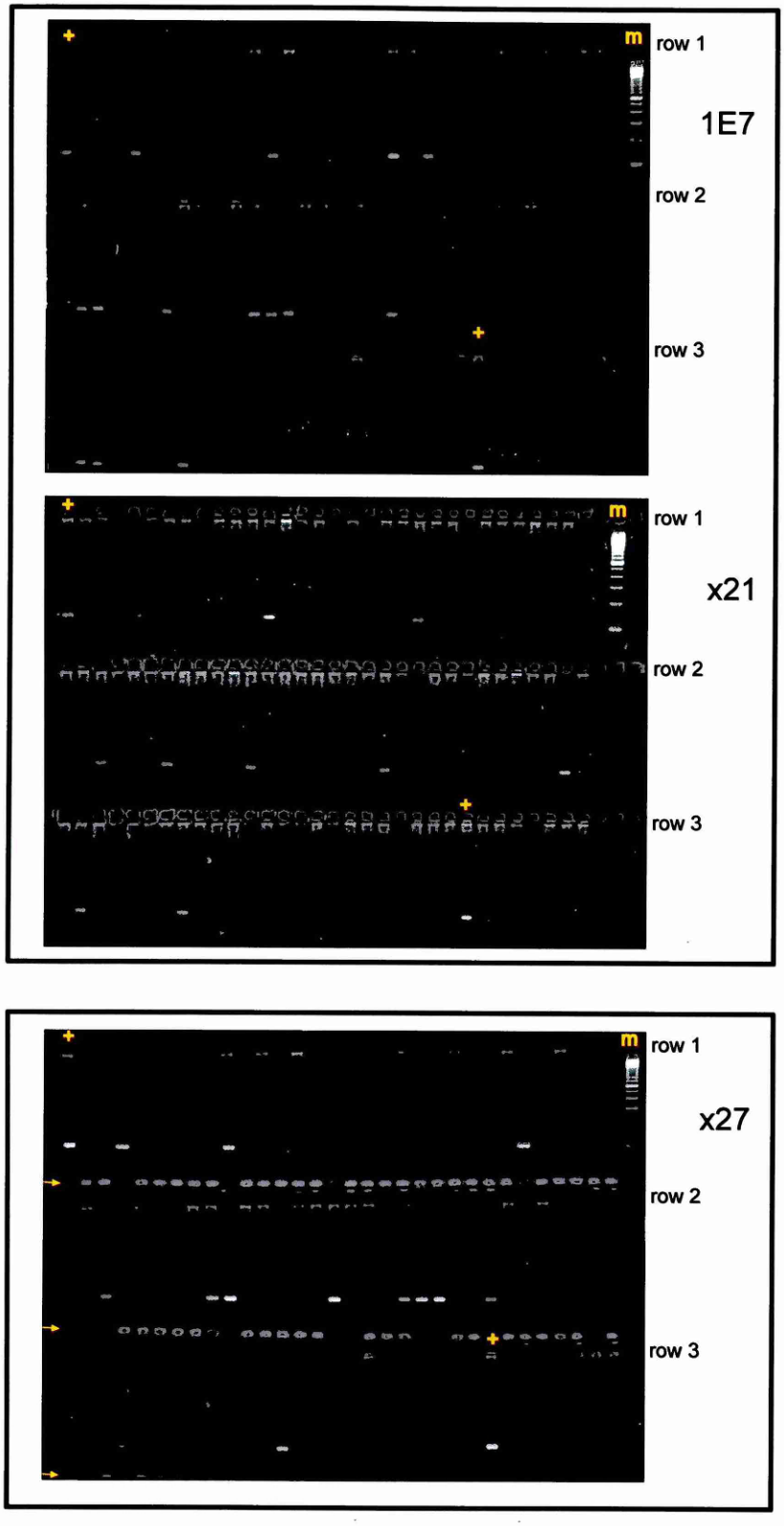


Figure 27 Photographs of ethidium bromide-stained agarose gels demonstrating WG-RH typing results.
There are three rows per gel, the DNA ladder (m) is present in the topmost row. The positive control (dog genomic DNA) is shown by +, arrows on the gel of x27 indicate primer-dimer. The retention patterns seen for markers 1E7 & x21 are similar and subsequent analysis showed that these markers were linked to each other and several other markers. The typing results for x27 have observable differences when compared to x21 & 1E7; analysis showed that this marker was not linked to these markers.

The retention of each marker was recorded from two separate experiments. If there were more than three discrepancies between these results, a further reaction was performed and the results combined. The final, consensus retention pattern was used to examine the linkage between all of the markers using the RHMAP programmes (version 3.0, Boehnke *et al.* 1991) according to the instructions provided in the online manual.

Initially, RH2PT analysis was used to determine markers linked with a lod score of 4.0 or greater. This provided pairwise lod scores between all of the typed markers enabling linkage groups to be established. It also provided pairwise distance estimates (in cR_{3000}) between linked markers.

The markers in each linkage group were ordered using MINBRK with the "branch and bound" method, a non-parametric method of analysis which is independent of assumptions about fragment retention and has been shown to give the most accurate orders of linked groups (RHMAP online manual, version 3.0, Boehnke *et al.* 1991).

The best candidate orders obtained by MINBRK were used with the MAXLIK programme to find the orders of markers within groups taking into account the fragment retention probabilities. The equal retention model was initially used for all groups. Subsequently, groups known to be located near the telomeres and centromere (from the cytogenetic data) were also tested using the centromeric model (Boehnke *et al.* 1991). The two results were compared but, in all cases, there was no difference in the order produced by the two models. The best order of markers within each group was at least 100 times more likely than the next best order.

In addition to markers from the small insert clones, markers were typed from additional sources, including five genes, one known pseudoautosomal marker (1E7 - N. Suter, thesis) and six cosmid library clones known to map to chromosome 1 (K338, H38, H32, H152, H233 & H254 Breen *et al.* 1999a and M. Breen personal communication).

The results from this analysis for a number of the X and 1-library linked markers have been used to construct a basic framework, shown below.

3.10.1 Chromosome X

Of 49 markers tested for WG-RH typing, 39 were optimised with the standard method, of which 29 were successfully typed. The Sanger protocol was used to optimise 20 markers of which 14 were successfully typed, including five previously typed with the standard method

included as comparisons. In total 38 markers (77%) gave results; the five markers typed with both protocols gave identical results. The typing results are given in appendix 5.

The two-point analysis of the chromosome X data produced eight linkage groups at lod 4.00, and nine unlinked markers (table 12).

Table 12 Linkage groups from RH2PT analysis of chromosome-X markers

The linkage groups shown were formed with lod scores ≥ 4.00

Linkage group	Markers
1	x21 1E7 x13 x31 x91 PDHA1 x78
2	x40 x33 x66 x69 x50
3	x57 x9 x30 x34
4	PGK1 x77 CHM
5	x68 F9 x58
6	x67 x74 x71
7	x64 x49
8	x1 x35
Unlinked markers	x12, x20, x27, x28, x47, x51, x60, F8c

The MINBRK and MAXLIK computer programmes were used to ascertain the best order and spacing of markers within the six groups containing three or more markers. These data are summarised in figure 28.

Linkage groups and distance estimates (cR_{3000}) were obtained from RH2PT, best order obtained from MINBRK and confirmed with MAXLIK, under equal retention model. X91 & PDHA1 are tightly linked markers unable to be separated with this panel; a higher resolution panel may be able to separate them.

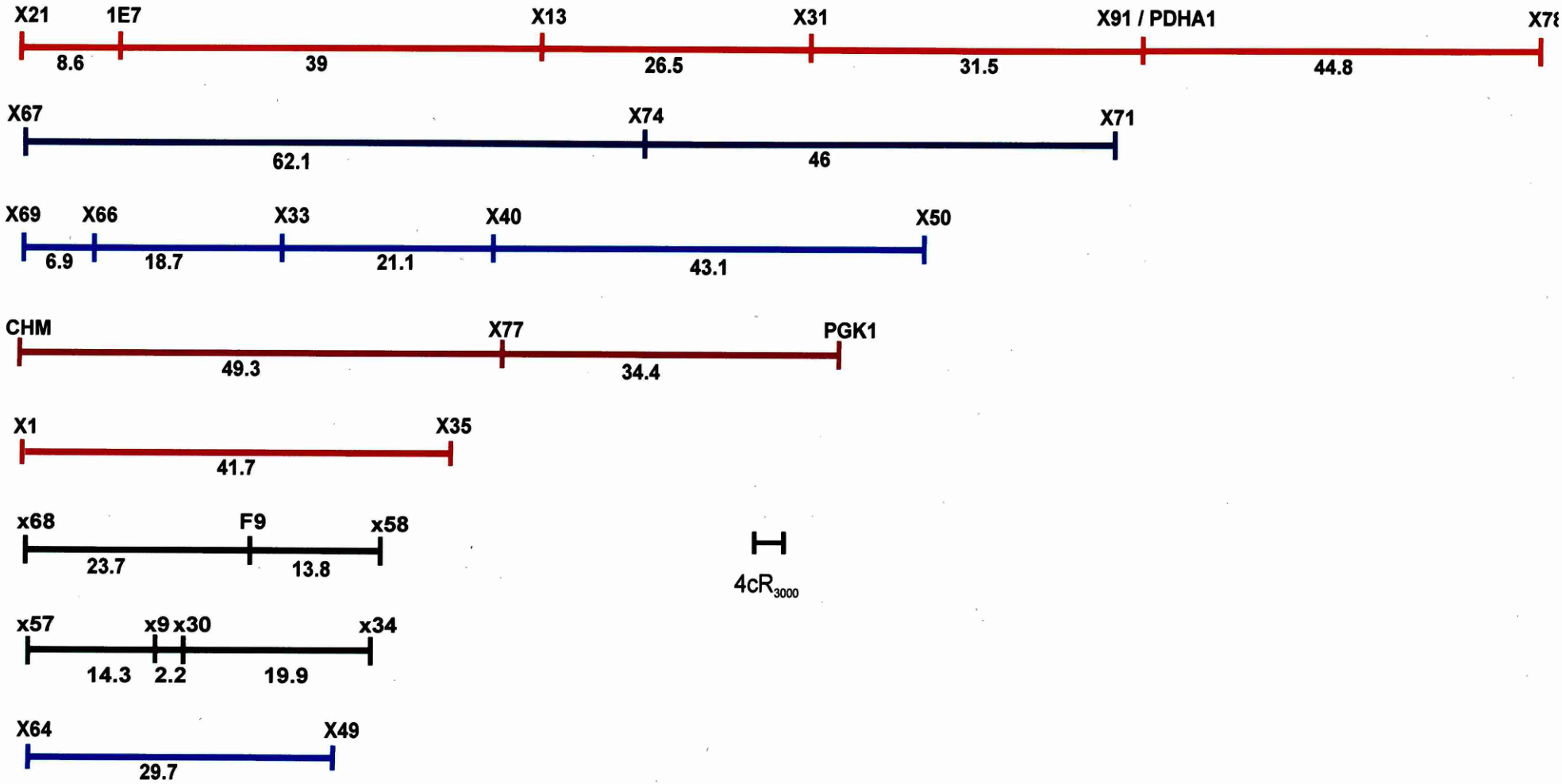


Figure 28 Orders of chromosome X markers obtained from two-point analysis using RHMAP.

Distances between markers are expressed in cR_{3000}

3.10.2 Chromosome 1

The chromosome 1 markers were only optimised using the standard method and of these, 18 out of 26 markers (70%) were successfully typed on the panel. The typing results are given in appendix 5.

The two-point analysis of the chromosome 1 data (table 13) produced five linkage groups at lod 4.00, and five unlinked markers. The MINBRK and MAXLIK computer programmes were used to ascertain the best order within the linkage groups that contained three or more markers, shown in figure 29.

Table 13 Linkage groups from RH2PT analysis of chromosome-1 markers

The linkage groups shown were formed with lod scores ≥ 4.00

Linkage group	Markers
1	1-60 K338 1-46 H38
2	1-32 1-27 H254
3	1-49 1-58
4	H152 H233
5	1-20 1-30
Unlinked markers	1-09, 1-11, 1-28, 1-42, H32

Linkage groups and distance estimates (cR_{3000}) were obtained from RH2PT, best order obtained from MINBRK and confirmed with MAXLIK, under equal retention model. H254 & 1-27 are tightly linked markers unable to be separated with this panel; a higher resolution panel may be able to separate them.

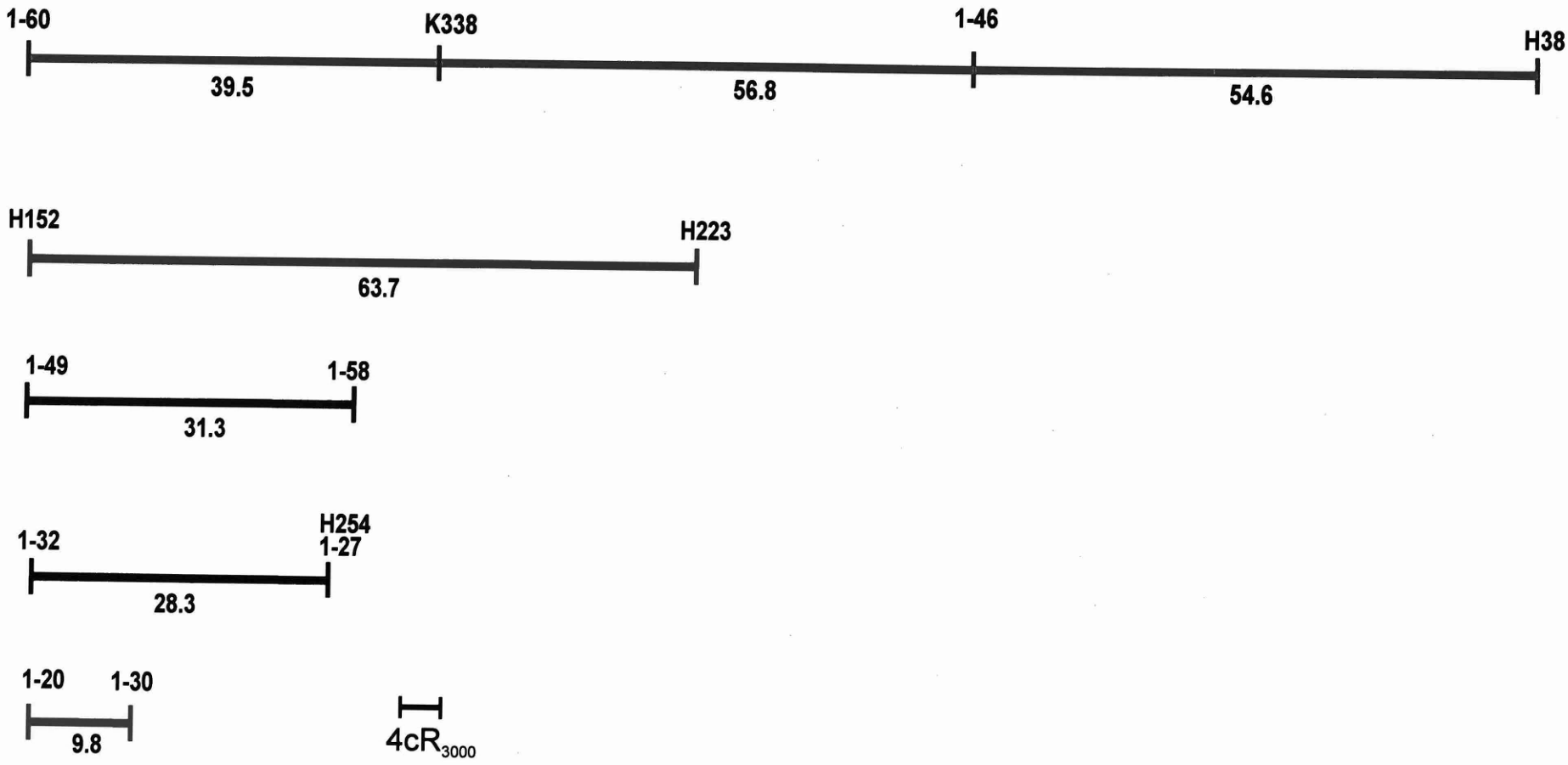


Figure 29 Orders of chromosome 1 markers obtained from two-point analysis using RHMAP.

Distances between markers are expressed in cR_{3000}

3.11 Integration of FISH, genetic linkage and radiation hybrid data from this study

Three mapping techniques were used in this study. FISH analysis was used to provide the absolute positions of cloned fragments on the chromosomes, typically to within 5 - 10 Mb, as indicated by the range over which signal was reproducibly seen. This technique provides absolute rather than relative locations and produces a direct, long-range map of the clones (and hence any markers derived from them) on the chromosomes, although at a low resolution.

Whole genome radiation hybrid (WG-RH) mapping provides information about the relative positions of closely spaced markers, but does not provide long-range information, nor can it give the absolute location or orientation of groups of markers. It does provide good short-range information but, like all segregation-based mapping methods, it is prone to errors at the ends of linkage groups (P. Deloukas, personal communication). The resolution of WG-RH mapping is dependent on the dose of radiation used to construct the panel, and hence the sizes of the fragments contained within the panel. All WG-RH maps will contain some misplaced markers because of the indirect nature of this method of analysis, including typing errors, incomplete data, PCR artefacts and inappropriate assumptions (such as using an inappropriate retention model or statistical method) made during analysis. Nevertheless, WG-RH analysis affords short-range mapping data of a far higher resolution than FISH.

The availability of FISH data in conjunction with WG-RH data for the same markers provide a powerful means of enhancing maps; FISH analysis provides a direct means of identifying the cytogenetic location of groups of markers that WG-RH mapping has ordered relative to each other, at high resolution.

Linkage data were available for chromosome X markers only and, because of the small size of the available pedigree and the fact that not all markers share informative meioses, were of limited power in this study. In the mouse (Dietrich *et al.* 1996), man (Dib *et al.* 1996) and the rat (Jacob, *et al.* 1995) scarcity in the number of microsatellites on chromosome X has been reported. In addition, the average heterozygosity of microsatellite markers on X in man (Dib *et al.* 1996) and mouse (Dietrich *et al.* 1996) has also been shown to be lower than for those located on the autosomes. Of the dog chromosome X-derived markers, less than half were polymorphic and informative in the reference families used, suggesting that there may be a

similar reduction in polymorphism of microsatellite-markers on dog chromosome X. Furthermore, the published genetic linkage maps of the dog (Neff *et al.* 1999, Mellersh *et al.* 2000) have very few microsatellite markers located on chromosome X (5 and 7 respectively), also suggesting that the dog chromosome X, as in man, mouse and rat, has fewer microsatellites located on it. Although limited, the linkage data from this study can assist in the integration of the FISH and WG-RH data with published maps, and can be included in future studies that use the Cornell families. Furthermore, if samples from the additional animals used in the published maps are made available, it may be possible to integrate the data produced in this study more fully.

3.11.1 Chromosome X

The FISH data identifies the physical location of 37 clones along the X chromosome, of which 35 were typed using the WG-RH panel and 29 were placed in a total of eight WG-RH linkage groups. Markers that were not mapped by one technique can still be allocated a position in the map by virtue of their inclusion in another map.

In all cases, the markers in each of the WG-RH linkage groups map by FISH to the same region of the chromosome, confirming that none of the markers have been significantly misplaced by either method. The relatively small size of some of the WG-RH groups, when compared to the range over which FISH signal was observed for the clones, means that some of these groups may only be tentatively oriented.

The first linkage group shown in figure 30, x21 - x78, could be oriented because two of the markers, x13 and x21 are well-resolved by FISH and also, x21, x13 and 1E7 are known to reside in the PAR, and therefore, to be at the distal end of Xp24. One marker in this group, x78 appears to reside at the end of the linkage group but it was physically mapped between x13 and x91. Since the ends of WG-RH linkage groups are prone to marker misplacement (P. Deloukas, personal communication), it is more likely that x78 has been accurately mapped by FISH and erroneously by WG-RH. The addition of more markers to the WG-RH map may resolve this anomaly since there would be more markers potentially available to be linked to x78 and which could remove x78 from the end of the linkage group.

The x71 – x67 linkage group was oriented by the FISH data of x71 and x67, since these markers are well resolved by the FISH data in this study.

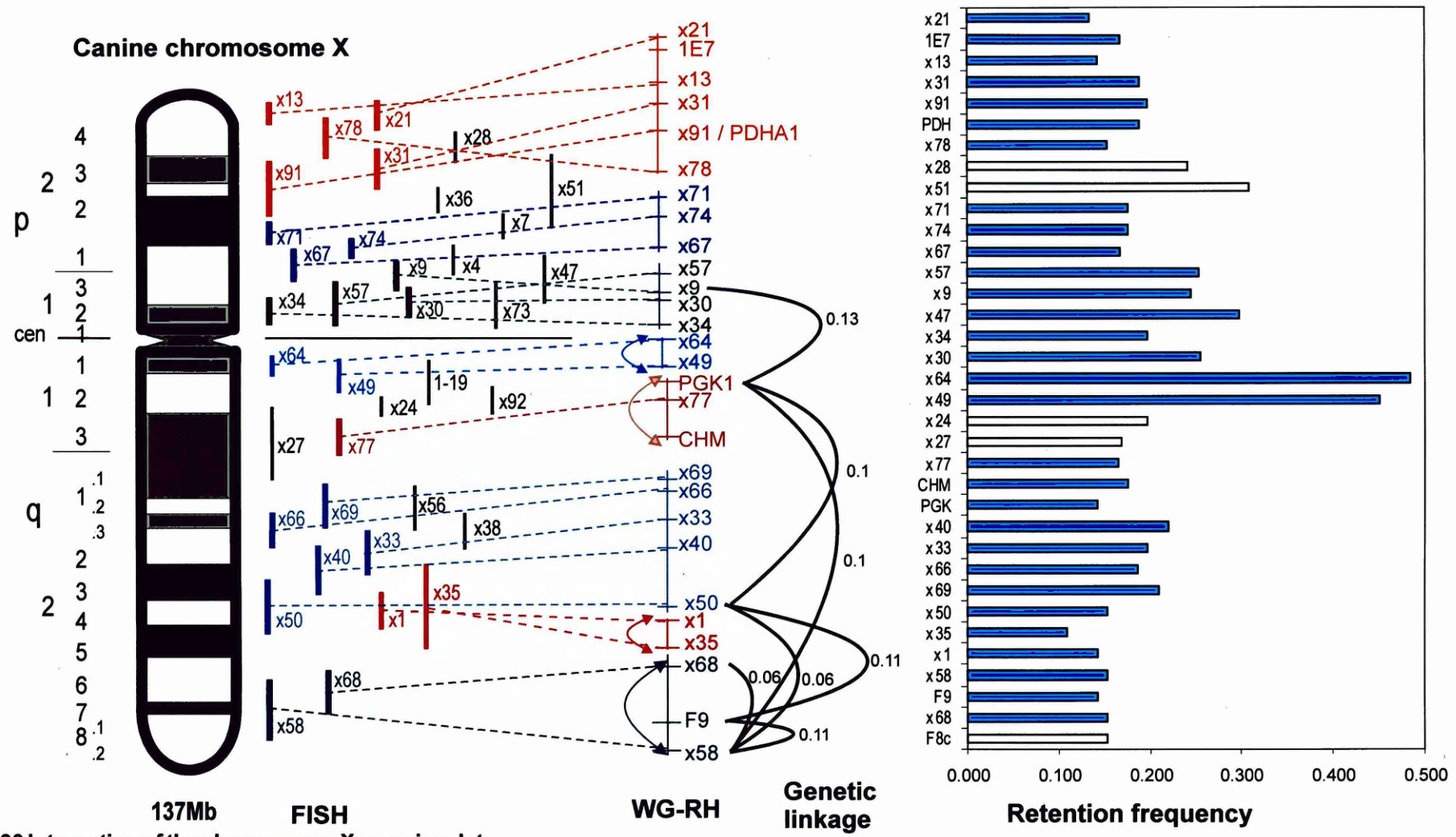


Figure 30 Integration of the chromosome X mapping data.

The ideogram is shown on the left. For markers mapped by FISH short vertical bars indicate the range over which signal was seen for each clone. Markers in WG-RH linkage groups are connected to the corresponding FISH marker by dashed lines. Double-ended arrows indicate WG-RH groups which cannot be oriented using the FISH data. Recombination fractions between significantly linked markers are indicated (Genetic linkage). The retention frequencies of markers in the WG-RH panel are shown on the right, the blue bars indicate linked markers & the grey, unlinked markers..

The orientation of the linkage group x57 – x34 was determined because x34 and x9 were well resolved by FISH in this study and x34 maps proximal to x9.

The linkage group containing x77 contains two type-1 markers, (PGK1 and CHM) and could not be oriented because only x77 has been physically assigned in this study. The physical assignment of these two gene markers by Deschênes *et al.* (1994) placed them in Xq, in agreement with the position of x77 by FISH in this study and the linkage data of Deschênes *et al.* (1994) suggested that PGK1 is proximal to CHM, as shown.

Linkage group x69 – x50 was oriented by the cytogenetic location of x50 and x69, which have no overlap by FISH analysis.

The final linkage group, x68 – x58 could not be oriented since the markers could not be fully resolved by FISH in this study. The WG-RH mapping data suggest that the type I marker, for clotting factor IX (F9) lies between the two, nearer to x58 than x68.

The retention frequencies, as shown by the bar graph, indicate a general trend for greater retention of fragments from the centromere, as has been observed on human chromosome X by Gorski *et al.* (1992). The retention frequencies accord with the mapping data and support the overall order of markers.

The linkage data, as already mentioned, is severely limited by the size of the pedigree available. The distances between the markers are likely to be overestimated, especially for the many markers that share few or no informative meioses. This is borne out by the theta values shown in figure 30, which are consistent neither internally nor with the FISH or WG-RH data. The linkage data therefore, will be of more value for future map integration once additional data is obtained.

3.11.2 Chromosome 1

The FISH data positions 23 small-insert clones along chromosome 1. Markers from 12 of these and an additional six cosmid-derived markers were also WG-RH mapped and are combined with the FISH data in figure 31.

As with the chromosome X data, all markers that are linked by WG-RH also co-localise by FISH indicating no significant mapping errors.

The linkage group H38 – 1-60 could not be oriented in this study, as the FISH resolution is not sufficient to resolve the four markers, however, K338 appears to be proximal to 1-60 suggesting that the orientation as shown in the figure is correct.

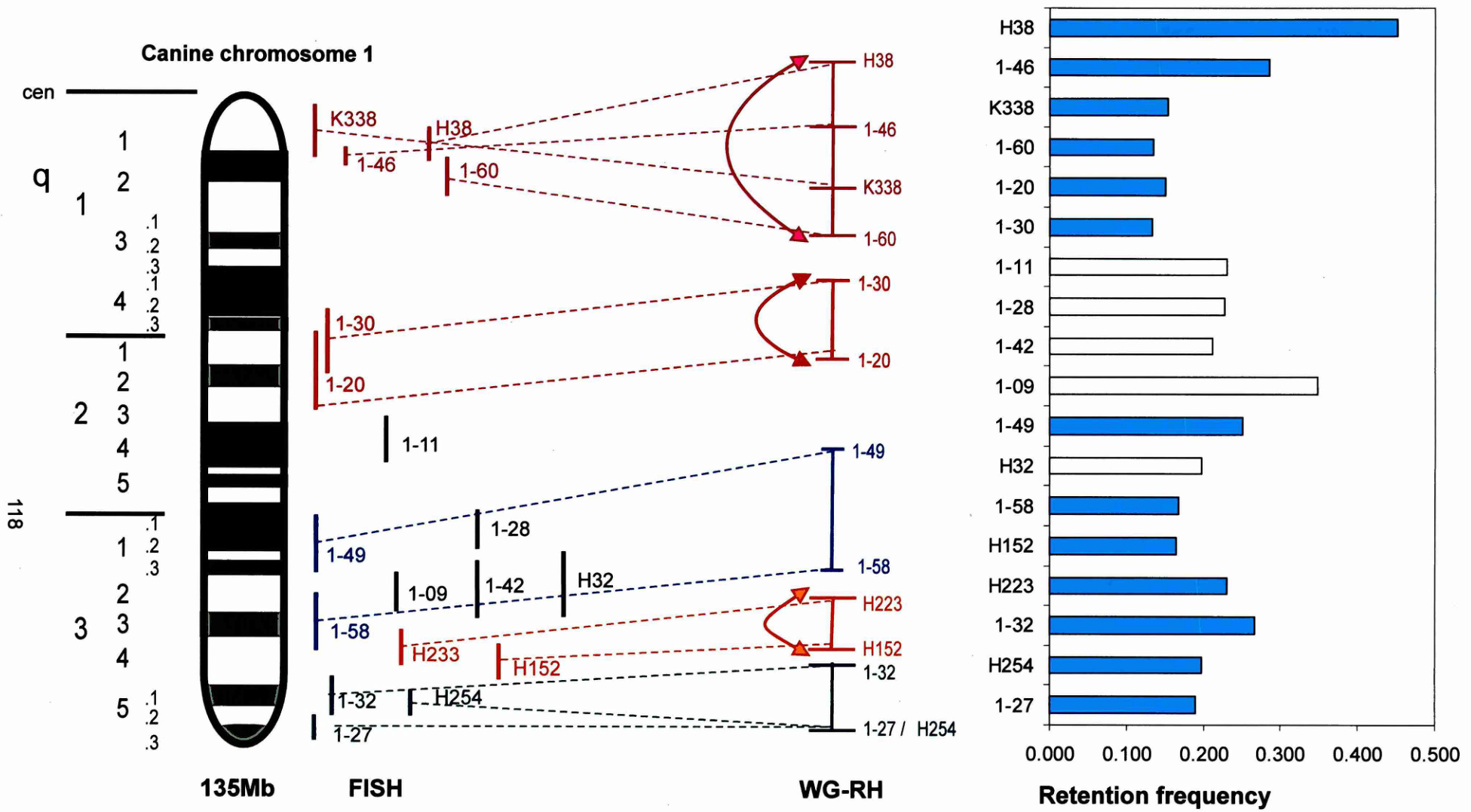


Figure 31 Integration of the chromosome 1 mapping data.
The ideogram is shown on the left. For markers mapped by FISH short vertical bars indicate the range over which signal was seen for each clone. Markers in WG-RH linkage groups are connected to the corresponding FISH marker by dashed lines. Double-ended arrows indicate WG-RH groups which cannot be oriented using the FISH data. The retention frequencies of markers in the WG-RH panel are shown on the right, the blue bars indicate linked markers & the grey, unlinked markers.

Markers 1-20 and 1-30 could not be oriented as they overlap by FISH analysis in this study.

The next group, 1-49 – 1-58 could be oriented since the two markers are well resolved by FISH in this study, with 1-49 being proximal to 1-58.

Markers H233 and H152 cannot be oriented, as they could not be fully resolved by FISH in this study.

The final group, 1-32 – 1-27, could be oriented, as 1-32 and 1-27 did not overlap significantly by FISH in this study.

The retention frequencies for the markers (shown in the bar graph) indicate an increased retention of centromeric fragments as indicated by the high value for H38 in 1q11, and the relatively high value for the adjacent marker, 1-46. This has also been noted on human autosomes, including chromosome 21 (Cox *et al.*, 1990, Burmeister *et al.* 1991) and chromosome 16 (Ceccherini *et al.* 1992). The remaining retention frequencies are similar except for marker 1-09, which appears to be exceptionally high. The markers 1-42 and H32 are located in the same region and have retention frequencies similar to the majority of other markers; it is therefore not clear why this marker has such a relatively high retention frequency.

None of the small insert, chromosome 1 markers were linkage mapped; however, the data produced from this study will assist in future mapping efforts. The microsatellite flanking sequences may be readily obtained and used in genetic linkage mapping. These markers, with known physical location and additional linkage data from WG-RH mapping, can then be integrated with the existing linkage data.

3.12 Integration of data from this study with existing maps

The information obtained in this study will be of increased use if it can be integrated with existing maps. None of the type I markers used in this study were mapped by FISH directly, although some information can be inferred from other maps and from the FISH data for PGK1 and CHM by Deschênes *et al.* (1996).

The WG-RH mapping data cannot be directly integrated with the radiation hybrid map of Priat *et al.* (1998) as the markers have been typed on different panels; this study on T72 (Research Genetics) and Priat *et al.* (1998) on WG-RH_{5000}. If markers were reciprocally typed

on both panels, then integration would be possible, as has been done with the human WG-RH maps using Genebridge 4 and the Stanford G3 panel (Deloukas *et al.* 1998).

The original DogMap linkage map cannot be directly integrated with data from this study as none of the genotyping was carried out fully on the DogMap families for this thesis. However, the DogMap data has been assimilated with the more recent genetic linkage maps of Mellersh *et al.* (1997) and Neff *et al.* (1999). The map of Neff *et al.*, (1999) contained details of additional information available electronically via the journal's web page (URL: www.genetics.org/supplemental); this included the genotyping data for the markers analysed in their study. This potentially enables the data generated from the small-insert chromosome X library to be integrated with the Neff map. However, the Neff map used a pedigree that contained the Cornell reference families and additional animals that were not available to the AHT - as shown in figure 32. Consequently, integration of the data from this study with the published data was difficult as there were frequently too few informative individuals in common between the datasets.

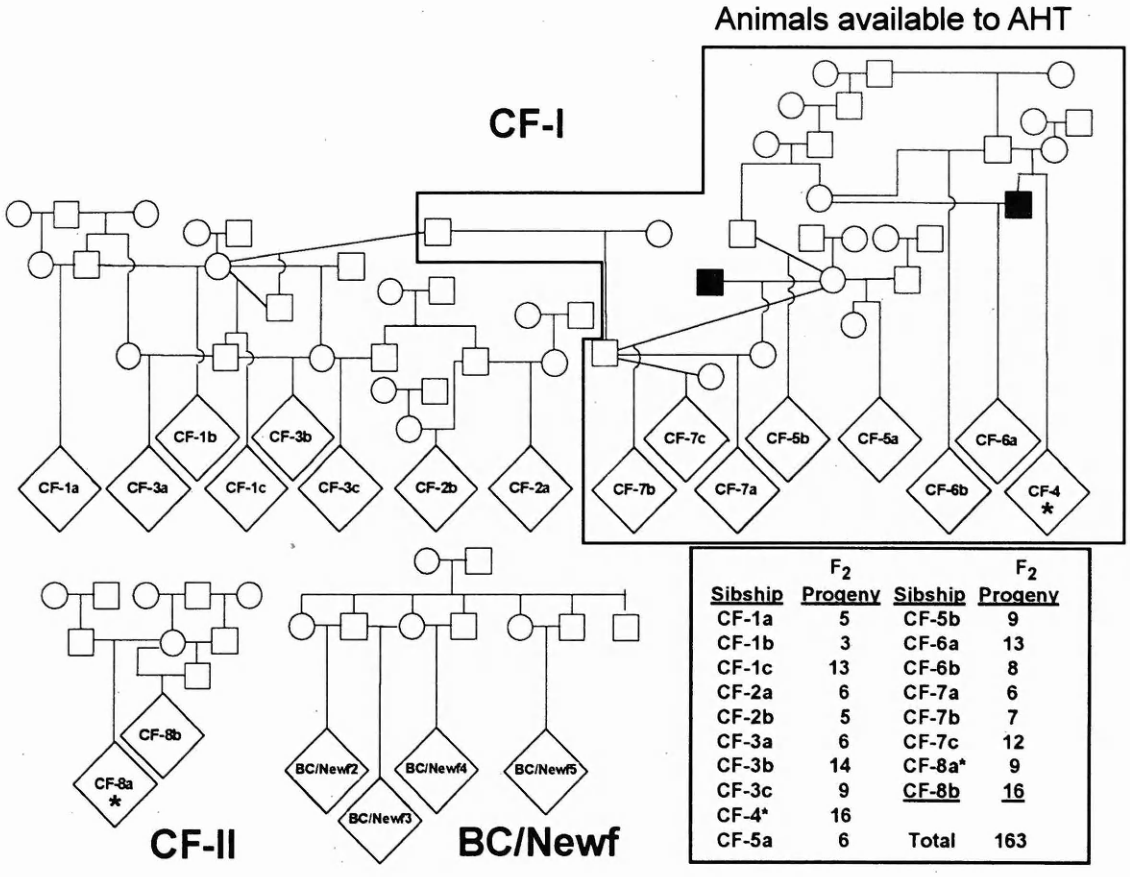
3.12.1 Chromosome X

The chromosome X map from Neff *et al.* (1999) contained two linked markers (AR and PGK) with three other markers (F8, COX.314 and CHM) mapped to chromosome X. When the typing data for these markers was examined, it was apparent that the linkage of these markers depended on data from animals unavailable to the AHT. It was therefore not possible to integrate the sex-linked data from the small-insert chromosome X library with the Neff map at this time.

Three of the markers from this study, x21, x28 and x51 showed patterns of inheritance suggesting they were pseudoautosomal or autosomal, i.e. some male members of the families were heterozygous for these markers.

Marker x21 is known to be pseudoautosomal from the FISH data and was informative in a small number of animals. Unsurprisingly therefore, it did not show significant linkage to any other markers, with those either from this study or the published data.

Marker x28 showed linkage to three of the markers in linkage group 7 of the Neff map (CXX.391: $\theta = 0.09$, lod 18.51; FH2348: $\theta = 0.22$, lod 5.71 and CXX.365: $\theta = 0.25$, lod 3.06) and to one marker previously unassigned to any linkage group, FH2394: $\theta = 0.25$, lod 3.59. The



Sibship	F_2 Progeny	Sibship	F_2 Progeny
CF-1a	5	CF-5b	9
CF-1b	3	CF-6a	13
CF-1c	13	CF-6b	8
CF-2a	6	CF-7a	6
CF-2b	5	CF-7b	7
CF-3a	6	CF-7c	12
CF-3b	14	CF-8a*	9
CF-3c	9	CF-8b	16
CF-4*	16		
CF-5a	6	Total	163

Figure 32 Taken from Neff *et al.* (1999), showing the Cornell families used to produce the second generation linkage map.
The animals available to the Animal Health Trust are indicated by the green box, (also illustrated in figure 3). The filled squares indicate the same individual represented twice for diagrammatic purposes. The asterisks indicate a pair of presumed identical twins.

primers from x28 amplify a polymorphic tetranucleotide, which does not appear to be sex-linked. The linkage data by itself cannot determine whether it is or not, but the FISH data for x28 localise the marker near the PAR (Xp24). Marker x21 is known to be pseudoautosomal but did not show linkage to x28 or the markers in linkage group 7. However, as x21 was informative in a small number of the animals, the ability to detect linkage for this marker was significantly reduced. The FISH analysis of x28 was carried out on female metaphase spreads so the opportunity to examine whether it was present on the Y chromosome was not available (it was not known that the marker could be pseudoautosomal at the time the FISH experiments were carried out and time constraints prevented this from being performed subsequently). Figure 33 shows the likely position of x28 in linkage group 7. Recent work has placed this linkage group into a syntenic group (syntenic group 8, Mellersh *et al.* 2000). It is therefore unlikely that the linkage data for x28 show a pseudoautosomal location for x28. Although FISH analysis of the clone mapped it to Xp24, it does not exclude the possibility that the clone was chimaeric, since the contaminating, autosomal fragment may have been too small to be reliably detected by FISH.

Marker x51 was linked to two of the markers in linkage group 11 (PEZ6: $\theta = 0.00$, lod 5.12 and VWF: $\theta = 0.00$, lod 5.12). The clone hybridises to Xp23; however, some signal was observed on several occasions to one of the smaller autosomes, possibly in the 21 - 38 range. The restriction digest of this clone produced only a single insert-derived fragment (data not shown); however, the FISH data showed that the clone hybridised to both chromosome X and an autosome (which was not able to be reliably identified by this study due to time constraints). This may be due to a pseudogene or repetitive sequence being present on either the X chromosome or autosome to which the clone can bind and the primers are amplifying the locus on the autosome. No known repetitive sequence was detected by BLAST for the sequences from x51, although this does not mean that this clone does not contain any repetitive elements. Figure 33 shows the likely position of to x51 in linkage group 11 of the Neff map. This linkage group includes the von Willebrand factor gene, which is known not to be located on chromosome X; therefore, the primers from x51 amplify a locus elsewhere in the genome. This linkage group has recently been assigned to chromosome 16 (Werner *et al.* 1999, Mellersh *et al.* 2000)

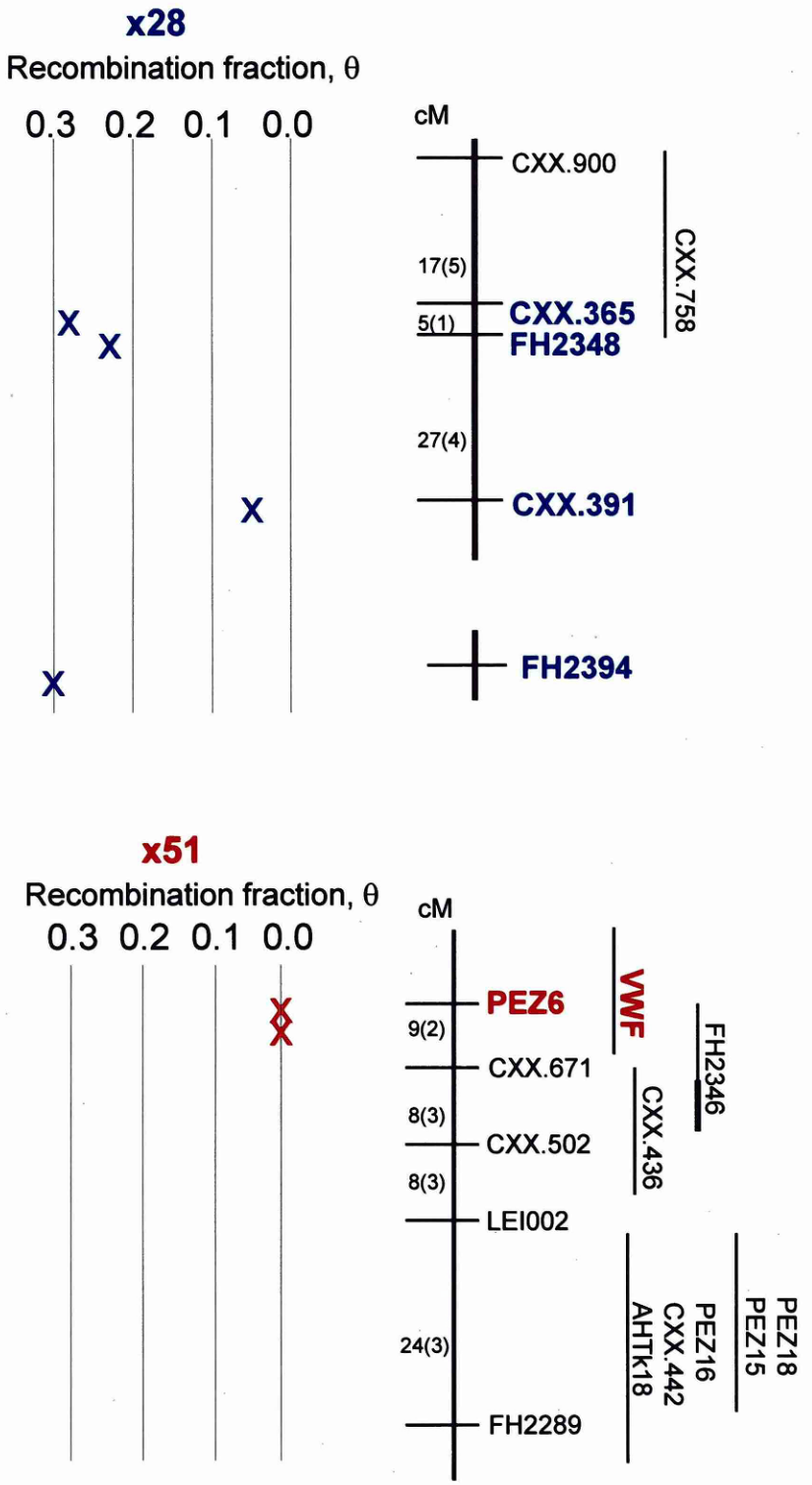


Figure 33 Linkage mapping results for markers x28 and x51 with data from Neff *et al.* (1999) for sex-averaged linkage groups 7 (top) and 11 (bottom). *Vertical dashed lines indicate recombination fractions, θ. Linked markers are shown in colour, "X" indicates the θ for each marker. These data add x28 & x51 to linkage groups 7 & 11. The data for x28 also add FH2394, a previously unassigned marker from the Neff map to linkage group 7.*

3.12.2 Chromosome 1

The marker K338 has been mapped by WG-RH mapping in this study and by FISH and genetic linkage mapping by colleagues at the Animal Health Trust. The linkage mapping carried out recently indicates that this marker is located in linkage group 1 of the Neff map (N. Holmes personal communication), in addition, it shows linkage to a marker unassigned to any published linkage group (FH2452) and therefore it assigns this marker to linkage group 1 and this linkage group to chromosome 1. The linkage data for K338 are as follows:

PGKAM: $\theta = 0.03$, lod 8.22

FH2016: $\theta = 0.06$, lod 11.52

FH2452: $\theta = 0.08$, lod 9.55

FH2313: $\theta = 0.11$, lod 9.04

CXX.673: $\theta = 0.14$, lod 3.84

CXX.424: $\theta = 0.22$, lod 3.53

The FISH analysis of K338 maps it to 1q11 – 12 (M. Breen personal communication), which places these markers near the centromeric end of chromosome 1. This linkage group and the location of K338 are shown in figure 34. The recent maps of Werner *et al.* (1999) and Mellersh *et al.* (2000) have confirmed the addition of FH2452 to this linkage group and have also assigned the linkage group to chromosome 1.

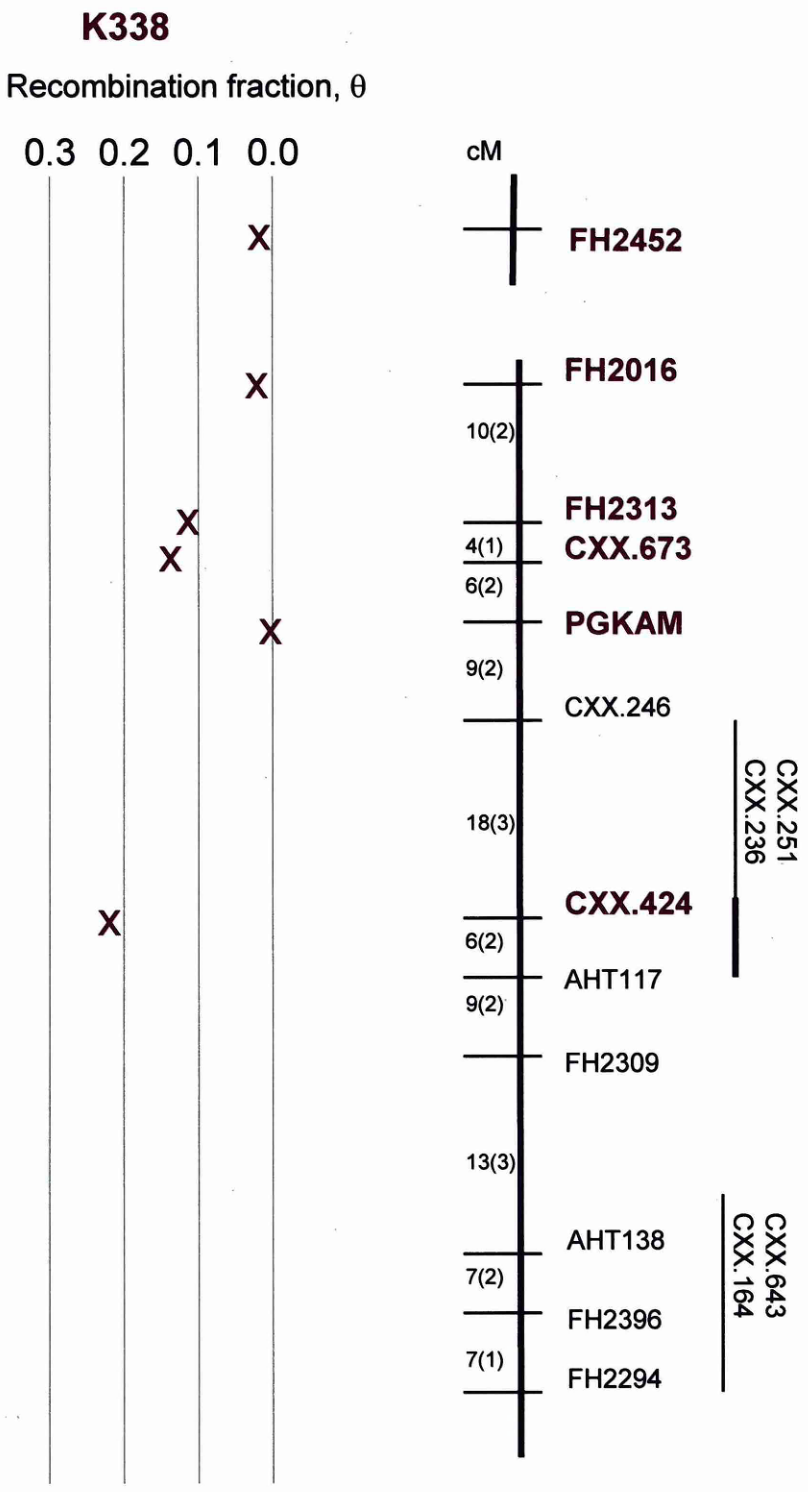


Figure 34 Linkage results for K338 with markers from the sex-averages linkage group 1 from Neff *et al.* (1999).

Vertical dashed lines indicate recombination fractions, θ. Linked markers are shown in colour, "X" indicates the θ for each marker. In addition to adding K338 and all the associated WG-RH & FISH data to linkage group 1, the results also add FH2542, a previously unassigned marker from the Neff map to this linkage group.

4 Discussion

When this study was begun, relatively little was known about the dog genome. In particular, little information existed for chromosome X, no microsatellite markers had been identified on chromosome X, and no linkage group had been assigned to either of the largest chromosomes, X and 1. It was apparent that no one technique would be adequate - FISH mapping lacks resolution, WG-RH mapping lacks contiguity (unless many markers are mapped) and linkage mapping was limited by the lack of a suitable reference pedigree. However, applying all three methods would provide a complementary and powerful set of data that could be integrated into a single map for each chromosome. The integration of all three techniques could be achieved by deriving and sequencing microsatellite-containing, chromosome-specific clones from the relevant chromosomes through flow sorting. Dog chromosomes X and 1 are well resolved in the dog flow karyotype, and so were particularly amenable to being isolated and sorted.

The combined approach to mapping these two chromosomes proved to be very effective. No one method is ideal but in combination they are very powerful and provide the best means of producing robust maps that could be integrated into future results.

The results produced by each of the methods will be discussed in turn, followed by a discussion of map integration and of the chromosomes themselves. Finally, future possibilities for dog genomics will be discussed.

4.1 Library construction, analysis and microsatellite distribution

Dog chromosomes X and 1 are well resolved in the dog flow karyotype, as can be seen in figure 5, section 3.1. They are the two largest chromosomes, are well separated from the rest and from each other in the flow karyotype and had both produced clean and specific paint probes in a previous study (Langford *et al.* 1996). However, the ability to generate large quantities of flow-sorted material from dogs is limited by the rapid degradation of the chromosomes when isolated (C. Langford personal communication). Despite this, sufficient material was produced to make the two libraries. It was unfortunate that the limited amount of

flow-sorted material meant that a proportion of it could not be spared to check the purity of the isolated chromosomes directly by chromosome painting. However, previous experience in making similar libraries at the Sanger Centre indicated that the libraries would be significantly enriched for the respective chromosomes (M. Ross personal communication). In addition, the subsequent FISH analysis of the clones derived form the libraries provided a means of assessing the purity.

The two chromosome-enriched libraries produced a similar percentage of recombinant clones, although the yield from the chromosome 1 library was slightly less than that for chromosome X. The yield of recombinants for both libraries is similar to that obtained by Longmire *et al.* (1993) for flow-sorted libraries of human chromosome 16 in Charon -40 (3.46×10^6) and sCos-1 (1.56×10^5). Although these vectors are phage and cosmid respectively, whereas the dog libraries in this study were prepared using the plasmid pBluescript, the methods employed by Longmire *et al.* (1993) are similar to those used in this study (they also obtained similar levels of purity in their libraries, with the phage library being 88% pure for human chromosome 16 and the cosmid library being 93% positive). Other flow-sorted chromosome specific libraries have been produced using linker-adapter PCR so are not directly comparable to this study (for example: Chang *et al.* 1992, Miller *et al.* 1992, Voojis *et al.* 1993, VanDevanter *et al.* 1994). However, from the data published by these groups, similar yields appear to have been obtained and therefore, the construction of the dog chromosome-enriched libraries may be considered to have worked well and yielded typical results.

The hybridisation-based method used to screen the clones for microsatellites of ≥ 12 repeats was successful, identifying 86 and 56 clones from the chromosome X and 1 libraries respectively. Sequencing identified microsatellites in the majority of the clones selected from the chromosome X-library (the remainder may have contained microsatellites that could not be identified by end-sequencing or which were unsuitable [i.e. were imperfect or compound repeats] for sequencing with the DRPs). Of these, the majority of microsatellites were of length $n \geq 12$; the average perfect repeat length was 16 and the maximum uninterrupted repeat length within compound repeats was 10.

Although the chromosome 1 clones were less well characterised due to time constraints, the end sequencing produced 15 microsatellites. These had an average perfect repeat length of 16 repeats and of 15 repeats for the longest uninterrupted stretch within compound repeats. The data from the X-library clones using the DRPs suggest that the

remaining chromosome 1 clones could be successfully sequenced and a similar yield of microsatellites obtained.

Work carried out previously to characterise the cosmid clones indicated that 328 of the 2116 clones examined (15%), were positive for the poly $(CA)_n$ probe (data not shown, work carried out prior to this study, by the author). The average cosmid insert size is 44kb (Stratagene) suggesting there is at least one $(CA)_n$ per 44/0.15 = 283kb along the dog genome, for $n \geq 12$ (each cosmid may contain more than one microsatellite, characterisation of some of these clones demonstrated this [N. Holmes, personal communication]).

The average insert sizes for the microsatellite-containing clones were estimated to be 7.5kb and 6kb for the chromosome X and 1-derived clones respectively. Screening for microsatellites (as for any other chosen sequence) will have biased the results to select the larger clones, and hence the average insert size for the libraries as a whole must be ≤7.5kb for chromosome X, and ≤6kb for chromosome 1. From these data, the minimum density of $(CA)_{n \geq 12}$ repeats in these libraries may be estimated: 1.42% of the chromosome X clones contained at least one $(CA)_{n \geq 12}$, implying that the mean spacing of such repeats on chromosome X is ≤7.5/0.0142, or at least one such repeat per 530kb. Similarly, for chromosome 1 (1.27% of clones found to contain a microsatellite; mean clone size ≤6kb) there is at least one such repeat every 470kb.

All of these estimates of microsatellite frequency are smaller than other published estimates (e.g. every 42kb; Rothuizen *et al.* 1994). The method used in this study for the identification of putative microsatellite-containing clones is optimised for detecting microsatellites with twelve or more repeats, whereas Rothuizen *et al.* (1994) obtained their estimate of microsatellite frequency from hybridisation with a $(CA)_{15}$ probe, under less stringent conditions; nor did they confirm their hybridisation results by sequencing. Their estimate of a microsatellite every 42kb may therefore be inaccurate, and the figure of 280kb based on cosmid clones (or <470 or <530kb based on the clones from flow-sorted chromosomes) may be a more accurate reflection of the density of $(CA)_{n \geq 12}$ in the dog genome. However, it should be noted that these results were obtained only from the microsatellite-containing clones, which will have larger inserts than the average for the rest of the library (selecting for the microsatellite-containing clones will bias the results to detect clones with large inserts). Estimates of dinucleotide frequency in the human genome suggest there is one every 30kb (Stallings *et al.*1991). However, this estimate is again based on hybridisation experiments and the

microsatellites were not characterised by sequencing. Thus, this estimate may reflect a broader range of dinucleotide repeat frequency, rather than the frequency of the longer, more polymorphic microsatellites as estimated in this study.

4.2 Sequence homologies

Chromosome X is an extremely well studied chromosome in many species, and hence a great deal of data have been deposited in the sequence databases. Chromosome X sequence is known to be conserved between species; therefore, more homologies were likely to be observed from the dog chromosome X sequences. The lack of significant matches from chromosome 1 markers to database entries is therefore not unexpected, especially since in looking for microsatellite-containing clones on chromosome 1 the sequences produced are most likely to not be associated with genes. The amount of sequence examined is very small compared to the total in the dog genome and to the amount of sequence in the database. It is important to remember that database searches are limited to the data deposited within them and although many sequences are present from human DNA, other mammals are less well represented. As the databases are constantly being updated, it is useful to periodically re-examine the data to find more homologies.

4.3 FISH

The FISH analysis of the chromosome-enriched clones yielded good results, despite the relatively small size of the clones used as probes. This study produced the first chromosome-X and -1 specific analyses in the dog. At the beginning of this study, no markers had been mapped to chromosomes X or 1 by FISH. During the course of the study, a number of microsatellite-containing clones were mapped to these chromosomes at the Animal Health Trust (Dickens *et al.* 1999). Several of these markers were included in this study to add more markers to the chromosome-specific maps and assist the integration of future maps with this work.

The dog karyotype is difficult to analyse as there are a relatively large number of chromosomes (n = 38), all the autosomes are acrocentric and show a gradual decrease in length. Chromosomes X and 1 are relatively straightforward to identify as they are the largest

chromosomes, and the X chromosome is metacentric. The Giemsa (and DAPI) banding patterns of the autosomes can be difficult to distinguish unless elongated metaphase preparations are consistently produced from which the subtle differences between the smaller chromosomes may be observed. The development of a set of chromosome-specific paints for the dog has made the task of identifying the individual chromosomes much more straightforward (Langford *et al.* 1996). In addition, the development of a set of microsatellite-containing cosmid markers on every chromosome, in combination with the paints, will enable unambiguous identification of each chromosome (Dickens *et al.* 1999, Breen *et al.* 1999a). The present study will complement this work by the inclusion of a number of the cosmid markers in addition to those developed from the chromosome-enriched libraries.

Dual-colour FISH has potential to resolve the order of markers by co-hybridising two or more clones, labelled with different colours and detected simultaneously. This was not routinely done in this study as the availability of WG-RH data provided much finer resolution and ordered the clones more accurately. However, the work carried out with the dogs known to be affected by or carrying the disease gene for muscular dystrophy used dual-colour FISH to show the relative position of the clones in the deleted region, x71 and x74, by using two clones that flanked this region, x9 and x36. The deletion was determined to be approximately 2Mb (Schatzberg *et al*, 1999 and M. Breen personal communication). The experimental work carried out with these animals indicated that the clone, x74, which had shown a homology to the human dystrophin gene sequence in GenBank, was indeed a true homologue. Several other putative genes (PHKA1, UTX, and ABC7) were located on chromosome X through these FISH experiments, although in these cases there is as yet no independent support for the homologies.

Constraints at the time the FISH experiments were carried out meant that not all the pseudoautosomal or possible pseudoautosomal clones could be examined on male spreads to confirm that they were located in this region. As the PAR is relatively small, it is not possible to say that all the clones tentatively assigned to it solely by FISH using female metaphase spreads are genuinely pseudoautosomal. Hybridisation of these clones to male metaphase spreads, or genotyping of the polymorphic markers on suitable animals, would determine this more accurately.

The X-derived clones were well distributed along the chromosome; although the region Xq27 – 28 was relatively sparse, no band was without a clone mapped to it. It was noted that

more clones were placed in Xp than in Xq, relative to its size (i.e. Xp = 1/3 X-length – 19 clones; Xq = 2/3 length - 18 clones). This may indicate the Xp was preferentially selected for during library construction, i.e. has higher density of restriction sites for *Hin*dIII, or that this area is denser in $(CA)_{n\geq12}$ repeats and therefore that more clones were detected from this region.

The FISH data for the chromosome 1 clones indicated that the middle region of the chromosome was under-represented by microsatellite-containing clones. The region 1q25 did not appear to have any clones located within it, although all other regions had at least one clone hybridising to them. The chromosome 1 clones showed some clustering in the regions 1q11-12 and 1q31-33, although fewer clones were unambiguously mapped to chromosome 1 in this study. Such a bias could arise from a difference in the distribution of *Hin*dIII sites or of microsatellites or by random chance.

The FISH analysis of the chromosome-enriched, microsatellite-containing clones provided the opportunity to examine the purity of the libraries and indicated that they are approximately 87% pure overall, with only 12 clones out of 92 found to hybridise to chromosomes other than those expected. There does not appear to be any pattern or significance to the contaminating chromosomes: they are distributed throughout the flow-karyotype, and are of a range of sizes and AT-compositions.

4.4 WG-RH linkage analysis

WG-RH analysis was successful for the majority of markers examined. Of the chromosome X markers, 29 were typed successfully using the standard protocol, with an additional eight markers optimised and typed using the Sanger Centre protocol producing 37 (68%) markers successfully WG-RH typed. Fewer markers from the chromosome 1 library were WG-RH mapped, but 70% of those tested (18 out of 26) were successfully mapped using the standard protocol alone.

Of the 17 X-chromosome markers which could not be typed, two failed because they did not amplify a clean PCR product, whilst the remaining 15 failed because co-amplification of hamster sequences could not be eliminated by further optimisation. In the case of chromosome 1, all of the ten failures were due to co-amplification of hamster DNA, although optimisation using the Sanger Centre protocol was not tested, and could perhaps have reduced this number.

Overall, therefore, approximately 30% of the markers from both libraries were unable to be mapped due to co-amplification of a hamster product. Rodents and carnivores are quite evolutionarily diverse and STS or microsatellite markers would not be expected to be conserved across species. However, many of the markers were found to contain a variety of different repetitive elements, including SINEs and LINES. SINEs have been shown to be conserved across a wide variety of species (Gilbert *et al.* 1999) and LINE elements have similarly been shown to be conserved across mammalian species (Bentolila *et al.* 1999). It may be that the presence of some of the sequence from repetitive elements within the PCR amplicon was sufficient to produce not only a canine product but one from the hamster DNA as well.

Although the marker density for both chromosomes was insufficient to link all of the markers, the majority of those typed fell into a small number of linkage groups. The availability of this WG-RH panel means that more markers can be typed in the future which will enhance the existing map and should join most of the present linkage groups together. Furthermore, the markers developed in this study can be typed on the recently developed WG-RH_{5000} panel, and thus permit the two maps to be integrated (see section 4.6 below). These two panels are complementary, since they have different resolutions; this has been the approach used in human genome WG-RH mapping using both the Genebridge 4, 3000-rad panel and the Stanford G3, 10,000-rad panel (Deloukas *et al.* 1998).

4.5 Genetic Linkage analysis

The linkage data produced for the chromosome X markers was limited by the fact that a suitable reference pedigree was not available until late in this project - the DogMap families available at the beginning were not ideal as they were only two-generation, phase-unknown pedigrees with limited amounts of DNA available. Although the Cornell reference families were more suited to genetic linkage mapping than the DogMap ones, the size and composition of the families available to the AHT was still rather restricted. The majority of the chromosome X-derived microsatellites generated in this study were typed successfully, although they were not all polymorphic in all the relevant animals. The number of published chromosome X markers is small - Neff *et al.* (1999) had only 5 sex-linked markers, and none shown to be pseudoautosomal. The linkage data from the flow-sorted chromosome-X markers could not be

integrated with the existing published data because the informative markers had too few individuals in common to show linkage. Nevertheless, the utility of the flow-sorted, chromosome-enriched microsatellite markers was demonstrated for the first time in the dog. Moreover, the genotyping data produced in this study is available to be integrated with future linkage mapping results.

4.6 The value of the integrated mapping approach

It was anticipated that the combination of mapping approaches, as well as making the best use of the resources, would prove a potent tool for dog genome mapping. This was demonstrated.

For both chromosomes, FISH provided valuable long-range mapping data and gave the absolute position, orientation and chromosomal assignments of the clones for the first time. The linkage maps of Lingaas *et al.* (1997), Mellersh *et al.* (1997) and Neff *et al.* (1999) did not directly assign microsatellite markers to any dog chromosomes. The assignments of a small number of linkage groups (four) to chromosomes in these studies was based on the presence of type I loci within the relevant linkage groups, which had been assigned to a chromosome in separate studies. The difficulty of accurately identifying individual dog autosomes has already been discussed, and the accuracy of chromosomal assignments used in these three linkage maps should be treated with caution. The application of chromosome-specific paints in conjunction with the markers of interest would permit a definitive assignment. Thomas *et al.* (1999) demonstrated the problems other workers have had in correctly assigning markers on the canine chromosomes. Previous studies had placed a number of type I markers on three different canine autosomes, chromosomes 3, 5 and 7 (Guevara-Fujita *et al.* 1996, Werner *et al.* 1997). Using reciprocal chromosome painting on dog, cat and human, Thomas *et al.* (1999) showed that these markers actually all mapped to just one chromosome, dog chromosome 5.

In order to produce the linkage map of chromosome X, Neff *et al.* (1999) used additional animals not available to the AHT at this time. Recent work by Werner *et al.* (1999) and Mellersh *et al.* (2000) has made progress in providing anchored and integrated dog genome maps. In order to produce these maps, both of these groups had to use animals in addition to the full Cornell Reference families to generate the maps. As with the map of Neff *et al.* (1999),

Mellersh *et al.* (2000) and Werner *et al.* (1999) found the Cornell reference pedigrees insufficiently powerful to make a whole genome linkage map of the dog. In this study, not even the full Cornell reference family was available to the AHT, making the task of linkage mapping on this family even more difficult. To fully integrate the data from this study with the recently published integrated maps, it would be necessary to obtain DNA from the additional animals and carry out more genotyping.

The WG-RH mapping data provides the local order of markers at high resolution, in contrast to the FISH map, which provides the absolute position of the markers. The availability of the WG-RH panel permitted non-polymorphic as well as polymorphic markers to be typed. It was therefore possible to type more markers using WG-RH than with genetic linkage mapping, and enabled type I and type II markers to be mapped simultaneously thus providing more data than the linkage mapping.

The radiation hybrid map produced in this study cannot be directly integrated with the published WG-RH map of Priat *et al.* (1998, WG-RH_{5000}) since this map was produced on a different panel. However, a subset of well-linked framework markers could be mapped on the WG-RH_{5000} panel and thus enable the two maps to be integrated. The higher resolution of the WG-RH_{5000} panel could be exploited to make a higher density map. It should be noted that the WG-RH_{5000} panel was constructed from a male animal, whereas the panel used in this study, T72, was derived from a female dog. This means that frequency of fragments in the WG-RH_{5000} panel from chromosome X will be half that of the panel used in this study. Integrating the two maps will require the analysis to be altered to allow for this - this can be done easily, as the programmes contain options for such a situation.

The data from the three mapping methods in this study may be consolidated to form an integrated map of each chromosome, as discussed in the following sections.

4.6.1 Chromosome X

The combination of FISH and WG-RH data allowed the production of an integrated map of both type I and type II loci on canine chromosome X for the first time, as shown in figure 30, section 3.11.1.

The integrated map includes six WG-RH linkage groups, three of which can be oriented by the FISH data. The FISH and WG-RH data are mutually consistent: in every case, markers that map to the same WG-RH linkage group map to the same region by FISH. This confirms the

accuracy of both methods and implies that the resulting map is robust within the limits of the resolution expected. This is significant because the majority of other maps rely on relatively limited data, with no external validation.

The linkage data were broadly consistent with the FISH and WG-RH (given the limitations already mentioned), and allow some integration with the other dog linkage maps. The recent integrated map of Mellersh *et al.* (2000) confirms the addition of the floating marker FH2349 of Neff *et al.* (1999) to the linkage group containing x28, and assigned to syntenic group 8 by Mellersh *et al.* (2000). Likewise, x51 shown in this study to be in the Neff map, linkage group 11, can now be assigned to chromosome 16 from the data of Mellersh *et al* (2000).

The markers shown to be located in the PAR of dog chromosome X by the FISH and linkage data indicate the size of this region. Three markers used in this study (the STS x13, the polymorphic marker x21 and 1E7 [N. Suter, PhD thesis, 1998]) were confirmed by FISH to be present on both the X and Y-chromosomes. The WG-RH data, in conjunction with the FISH data, place these markers in the distal region of Xp24. The WG-RH linkage group also contains three STS markers and a type I marker, PDHA1. There is no evidence from the FISH or genotyping data (where applicable) to say whether these three markers are pseudoautosomal in the dog; however, data from mouse, man and rat indicate the PDHA1 is not located in the PAR of these species. It is unlikely, therefore, that PDHA1 is located in the dog PAR. Further FISH analysis of the clones proximal to this region could help determine the limits of the dog PAR. In man, the largest PAR has been estimated to be 2.6Mb long (Rappold, 1993). From the ideogram of Reimann *et al.* (1996), Xp24 in the dog covers approximately 12% of the chromosome. Langford *et al.* (1996) estimated that dog chromosome X is 135Mb long, so Xp24 can be estimated to be approximately 16Mb long. It is unlikely that the dog PAR occupies all of Xp24: the PARs of other mammals studied to date appear to be of similar sizes to the 2.6Mb in man (Graves *et al.* 1998a, 1998b). In order to define the dog PAR in this region, it will be necessary to map genes known to be located both within and beyond this region. The availability of a relatively high number of markers mapped to dog Xp in this study will facilitate future characterisation of the dog PAR as they can be used in dual-colour FISH on male metaphase spreads to determine which of them are pseudoautosomal. The microsatellites contained within the majority of these clones can also be used in genotyping studies to look for male heterozygotes, which indicate that the markers are located in the PAR. In addition, the

canine BAC library (Li *et al.* 1999) could be screened with type-I loci known to map to the PAR in other species, the positive BACs can be FISH mapped, sequenced and STS or homologous markers derived form these, and used to make denser maps of this region.

The work presented in this thesis provides the first localisation of a number of genes or putative genes on dog chromosome X. Sequence similarities between clones developed and mapped in this study and genes known to be located on chromosome X in other species suggest tentative locations for PDHA1, UTX, ABC7, PHKA1, F9 and F8c. In addition, the location of the DMD gene in dog was confirmed experimentally (Schatzberg *et al.* 1999). The inclusion of PGK1 and CHM in this study provides the opportunity to integrate these data with those already published for these markers (Deschênes *et al.*, 1994, Mellersh *et al.*, 1997 and Priat *et al.* 1998). The validity of the putative gene homologies requires confirmation experimentally. If confirmed, they provide data of interest to researchers working on Menkes disease (ABC7), haemolytic anaemia (PGK1), tetratricopeptide repeat proteins (UTX), muscle glycogenosis (PHKA1), pyruvate dehydrogenase deficiency (PDHA1), haemophilia A (F8c) and haemophilia B (F9) (OMIM).

The presence of genes and homologues in this study allows the comparison of the dog chromosome X with those of other species; this will be discussed in more detail in 4.9.

4.6.2 Chromosome 1

The data from chromosome 1 provide a set of markers known to be located on the largest dog autosome. The lack of sequence homologies for chromosome 1 markers was predicted and future work could use sequence from more clones from this library to try to identify genes on chromosome 1. As no genes have been directly mapped to chromosome 1 to date, the cytogenetically assigned clones produced in this study are a valuable resource if used in this way. This would also provide more information about syntenic relationships between dog chromosome 1 and other mammals. Breen *et al.* (1999) used reciprocal ZOO-FISH to paint both human and dog chromosomes. They found that four human chromosomes (18, 6, 9 and 19) mapped to canine chromosome 1. Gene primers from the syntenic human or mouse chromosome regions could be used to screen the chromosome 1-enriched library; although it should be noted that the number of clones produced was 4416 and the size of chromosome 1 is 135Mb, so even if the average insert size of 6kb is not an overestimate, the small-insert chromosome-1 library only represents approximately 20% of dog chromosome 1.

As in the case of the X-chromosome, the integration of FISH and WG-RH data was a powerful tool. Similarly, both methods agreed in the co-localisation of groups of markers indicating that both techniques give reliable results, with no major misplacements of markers.

The radiation hybrid mapping produced a framework on which more markers may be positioned in future work. In combination with the FISH data, this provides a useful resource for the integration of the existing linkage data of Neff *et al.* (1999). The marker K338 has been mapped in this study and by colleagues at the Animal Health Trust. The linkage mapping carried out recently indicates that this marker is located in linkage group 1 of Neff *et al.* (1999) (N. Holmes, personal communication) and therefore it assigns linkage group 1 to chromosome 1. The marker appears to be located between CXX.673 and PGKAM, which would place these markers at the centromeric end of chromosome 1 since K338 has been mapped by FISH to 1q11 – 12 (M. Breen personal communication) This linkage group and the location of K338 are shown in figure 34. The recent map of Werner *et al.* (1999) used the FISH location of marker ZuBeCa2 (Schläpfer *et al.* 1998), to assign linkage group 1 to chromosome 1. Mellersh *et al.* (2000) integrated and extended both the linkage data of Neff *et al.* (1999) and the WG-RH_{5000} data of Priat *et al* (1998) with the markers anchored to chromosomes by Werner *et al.* (1999) to produce a comprehensive map of the dog genome. The integrated map of Mellersh *et al* (2000) confirmed the placement of the floating marker, FH2452 in the Neff map, into the chromosome 1 linkage group. The data produced in this study will add to the integrated map of chromosome 1 by providing a set of markers (most of which contain a microsatellite) with known cytogenetic locations and T72 WG-RH mapping data.

4.7 Comparative genomics of chromosome X

The mapping of several genes (and gene homologues) to chromosome X in this study make it possible to examine the syntenic relationships between dog and other species. Figure 35 illustrates the syntenic relationships between dog and eight other mammalian species. The genes illustrated have been mapped in the dog by a variety of methods. The clones from this study with database similarities to genes in other species have been included, but they must be considered putative candidates, as it has not been confirmed that they are true homologues. The information for the other mammalian species is derived from Internet sources or journals as indicated, and is from a combination of linkage and physical mapping data.

PDHA1 is located within RHMAP linkage group 1, which contains some known PAR markers. The putative physical location of this marker has been derived from the RHMAP data of neighbouring markers and their cytogenetic position. The primers for this gene are taken from Jiang *et al.* 1998 but were not mapped by this group.

DMD is positioned from FISH analysis of clone x74 located to Xp21, via the sequence similarity of this clone to the GenBank entry for human dystrophin. FISH of the clone in dogs affected by and carrying the disease muscular dystrophy (Schatzberg *et al.* 1999) determined further evidence that this clone is part of the true dog homologue of dystrophin.

FISH analysis of clone x57, which is located in Xp 11 – 12, tentatively positions UTX, through the sequence similarity of x57 to GenBank entries for this gene in man and mouse. This marker has also been mapped using the dog WG-RH panel and located within RHMAP linkage group 3.

FISH analysis of clone x92 tentatively positions ABC7, which is located in Xq12, through sequence similarity to GenBank entries for this gene in man and mouse.

FISH analysis of clone x24, which is located in Xp 11 – 12, tentatively positions PHKA1, through the sequence similarity of x24 to GenBank entries for this gene in rabbit, man and mouse.

F9 was mapped in this study and found to be located within RHMAP linkage group 5, which contains x58 and x68. The putative physical location of F9 is derived from the cytogenetic position of these two clones.

F8c was mapped in this study and tentatively found to be linked to x68 by RHMAPPER analysis of the WG-RH data. (The WG-RH results were also mapped using RHMAPPER at the Sanger centre (McCarthy & Soderlund, 1998). However, the small number of markers typed meant that this was unsuccessful in all but one case, where two markers (x68 and F8c) appeared to be linked; these were not shown to be linked by RHMAP analysis. In view of the limited dataset this method of analysis was not pursued.) It was also mapped to chromosome X by Mellersh *et al.* 1997. The physical location of this gene is suggested from the cytogenetic position of the linked marker, x68.

The genes CSF2RA, ANT3, STS were assigned to Xp/Y PAR1 using FISH by Toder *et al.* 1997.

PGK1 was mapped in this study using WG-RH and linkage mapping. Furthermore, AR, PGK1 and CHM were shown to be linked by Deschênes *et al.* 1994, and PGK1 and CHM were also physically mapped by FISH to proximal Xq. PGK1 and AR were linkage mapped by Mellersh *et al.* 1997. Priat et al. 1998, also mapped PGK1 and CHM by WG-RH mapping. The physical location of these markers has been estimated from all these data sources.

The data for the genes mapped in common were obtained from the sources indicated in table 14. These genes represent a small number of those mapped in these species but are common to many of the species described. The relative positions can be seen and the rearrangement of genes is observable between these eight species. The most comprehensive mapping has been carried out in man and mice. Comparison of dog with man indicates that the putative positions for the dog genes are similar to those in human. The five genes AR, ABC7, PHKA1, PGK1 and CHM appear to map to the same area of dog chromosome X although the order of these markers has not been determined in the dog. If x57 is a genuine homologue of UTX, then its position in the dog is similar to that in man. Likewise, if x92 is homologous with ABC7 then it would appear to be located in a comparable position to man. The order of genes in mice is rearranged compared to man and, it would appear, to dog. The other species shown in figure 35 have fewer genes mapped in common, but there would appear to be a range of gene positional rearrangements between all the species described. Further work in the dog is required before a definitive comparative map may be made, but the data from this study indicate that the dog chromosome X is far more similar to that of man, with respect to these genes, than those of the other species shown. This has obvious implications for the use of the dog as an animal model of human disease and indeed, the dog is already used as a model for

immunodeficiency disease (Deschênes *et al.* 1994) and muscular dystrophy (Sharp *et al.* 1992, Henthorn *et al.* 1994, Schatzberg *et al.* 1999).

If the putative genes described in this study are confirmed to be true homologues in the dog, then figure 35 illustrates a striking conservation of synteny between the dog and human X-chromosomes. It can be seen that there appear to be no major inversions of the dog genes examined, when compared with human. When compared to rat, two markers, PDHA1 and STS appear to be inverted with respect to each other, but the other four genes appear to be in the same orientation. The sheep similarly has a rearrangement of PDHA1 when compared with the dog, as in the sheep it is located in Xq not Xp, as in the dog. The other five genes appear to be syntenically conserved between sheep and dog. The goat has fewer genes mapped in common, but the F9 gene in the goat appears to be inverted relative to the dog but the other three genes appear conserved in order. Cat appears to be syntenically conserved for the three genes in common with dog. Cattle and pig have only two genes mapped in common with the dog. The order of PGK1 and DMD appears to be inverted in the cattle genome, when compared to the dog. It also appears that AR and PGK1 are inverted in the pig, although the assignment of AR in dog was not determined cytogenetically, and so may be in agreement with pig.

Only a relatively small number of X-linked genes are described in this study and more experimental work needs to be performed to determine whether these potential syntenic relationships are genuine. Nevertheless, in comparison to other model organisms such as rat and mouse, it demonstrates the potential utility of the dog as a model organism and indicates that it merits further study of the dog X-chromosome.

The comparisons shown contain data from a number of mammalian orders. The differences between rat and mouse (both members of the Rodentiae) and the similarities between man and dog (different orders) show that evolutionary closeness is not always a valid predictor of the syntenic relationships between species.

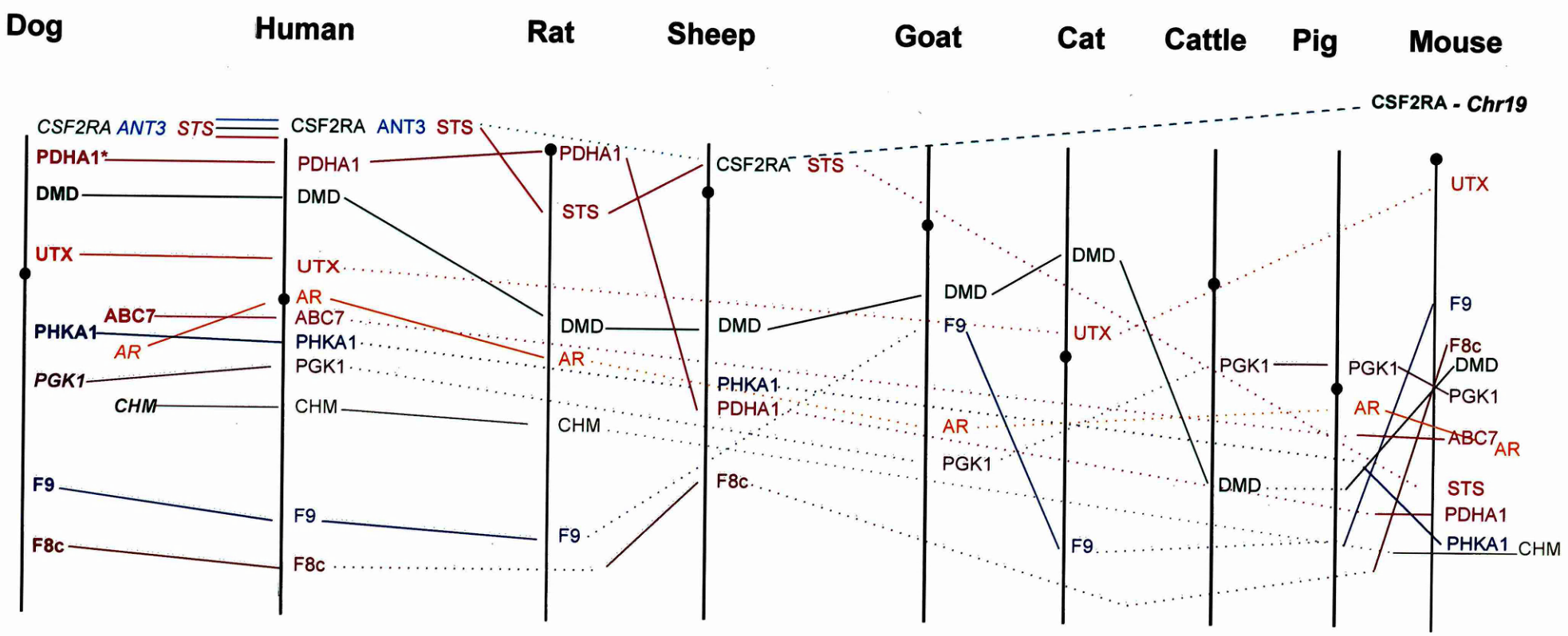
Table 14 Sources of information for genes located on chromosome X in eight species of mammal

Mouse	F9, DMD, F8c, PGK1, UTX, ABC7, PDHA1, PHKA1, CHM	Mouse genome database, URL - http://www.informatics.jax.org/menus/homology_menu.shtml
	AR	Millwood *et al.* 1997
	CSF2RA	Disteche *et al.* 1992
Rat	PDHA1, STS, DMD, AR, CHM, F9	Millwood *et al.* 1994 & Ratmap URL - http://ratmap.gen.gu.se
Cat	DMD, UTX, F9	Murphy *et al.* 1999
Pig	PGK1, AR	Pigmap URL - http://www.ri.bbsrc.ac.uk/pigmap/pig_genome_mapping.html
Cattle	PGK1, DMD, F9	Barendse *et al.* 1994 & Bovmap URL - http://www.ri.bbsrc.ac.uk/cgi-bin/mapviewer?species=cattle
Goat	DMD, F9, AR, PGK1	Piumi *et al.* 1998 & Goatmap URL - http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/goatmap/Goatmap/main.pl
Sheep	DMD, PHKA1, PDHA1, F8c	Galloway *et al.* 1996 & Sheepmap URL - http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/browser.sh?species=sheep
	CSF2RA, STS	Toder *et al.* 1997
Human	PDHA1, DMD, UTX, AR, ABC7, PHKA1, PGK1, CHM, F9, F8c	Genome Database (GDB) URL - http://www.gdb.org/gdbreports/GeneByChromosome.X.alpha.html

Figure 35 Conservation of genes on the X chromosome of eight mammalian species.

Vertical lines represent chromosome X in a number of mammals (not to scale); filled circle indicates relative position of centromere. Information about the relative locations of the genes was derived from Internet sources or journals and is a mixture of linkage and physical mapping data. The genes illustrated have all been mapped in the dog by a variety of methods. The clones from this study with database similarities to genes in other species have been included, but they must be considered putative candidates, as it has not been confirmed that they are true homologues. Dog genes located by this study are indicated in ***bold,*** *those by other groups are indicated in italics, those shared between this study and others are in* ***bold italics****. Solid lines connect genes mapped in each species, dashed lines indicate the gene has not been typed in the species that the lines cross.*

The figure illustrates the rearrangements in relative gene order that have occurred in mammalian evolution.


Dog
Human
Rat
Sheep
Goat
Cat
Cattle
Pig
Mouse
CSF2RA - Chr19
CSF2RA ANT3 STS
PDHA1*
DMD
UTX
ABC7
PHKA1
AR
PGK1
CHM
F9
F8c
CSF2RA ANT3 STS
PDHA1
DMD
UTX
AR
ABC7
PHKA1
PGK1
CHM
F9
F8c
PDHA1
STS
DMD
AR
CHM
F9
CSF2RA STS
DMD
PHKA1
PDHA1
F8c
DMD
F9
AR
PGK1
DMD
UTX
F9
PGK1
DMD
PGK1
AR
ABC7 AR
STS
PDHA1
PHKA1 CHM
UTX
F9
F8c
DMD
PGK1

5 Further work

This study has generated the first robust, integrated maps of dog chromosomes X and 1, but it also forms the basis for future work.

The clones, their sequences and the microsatellites represent a unique resource available for further dog genome mapping. The data from this study can be integrated into other maps by analysing the markers from the chromosome-enriched libraries on the reference resources used to make other maps. The genotyping and WG-RH data produced in this study can be incorporated into future studies that use the same mapping resources, i.e. the WG-RH_{3000} panel and the Cornell reference families.

This study has also demonstrated the feasibility of obtaining robust and versatile chromosome-specific markers from flow-sorted material. It should be possible to extend this approach to those other canine chromosomes that are well resolved in the flow karyotype. For those autosomes which cannot be uniquely resolved in this way (Langford *et al.* 1996), it should still be possible to produce libraries which target small groups of chromosomes, when a whole-genome approach might be too labour intensive, expensive or time-consuming.

Finally, it is hoped that the work described here has demonstrated the increased value that results from combining several distinct mapping techniques linked by a common set of markers. Such integration best exploits the available resources, produces maps of greater robustness than would be possible using a single technique, and could profitably be applied to other chromosomes of dog and other species.

6 Conclusions

The method used to create two chromosome-enriched libraries was successful. The majority of the markers generated were specific to the chromosome of interest. Microsatellite markers were generated for chromosome X thus addressing the dearth of markers available at this time on the chromosome in the dog. The additional data from gene markers and other libraries provides a useful means of integration and comparison with data from other species. The linkage data derived from the markers on chromosome X provide a framework into which other markers may be added.

Similarly, the radiation hybrid mapping data for chromosome X and 1 will permit additional markers to be positioned within a useful framework. The cytogenetic data will allow the absolute position of any marker associated with either the radiation hybrid or genetic maps to be determined.

The use of a common pool of markers in all three mapping approaches facilitates the integration of new data with each map and thus relates the physical and genetic maps. A set of polymorphic markers from chromosome X have been characterised and mapped. A set of STS markers for chromosome 1 have likewise been characterised and mapped. Several previously published markers have been mapped in conjunction with the chromosome-specific markers and can therefore be positioned on the relevant chromosome. Several thousand clones have been produced, the vast majority of which will be specific for the two chromosomes in this study and can be employed in future work.

The putative or actual locations of nine genes on dog chromosome X have been described: PDHA1, DMD, UTX, ABC7, PHKA1, PGK1, CHM, F9 and F8c. Two of these have been mapped previously and the data from this study enhance the knowledge of the locations of these genes on dog chromosome X. The syntenic relationships between the dog X chromosome and the X chromosomes of other mammalian species have been investigated and indicate that the dog may be the species with the greatest conservation of synteny on chromosome X when compared to the human X chromosome.

Dog genomics is still in its early stages. However the data presented here provide a valuable enrichment of the knowledge of two dog chromosomes. The value of the dog as a model organism has been enhanced by the syntenic relationships between dog and man by

providing information about genes inaccessible in other species. This emphasises the usefulness of continued study of the dog genome and has additionally demonstrated the power of an integrated mapping approach.

Appendix 1

General solutions

All reagents were molecular biology grade from Sigma unless otherwise stated. All water was MilliQ grade.

Agarose gel loading buffer

62% glycerol v/v

125mM Na_2EDTA

62mM Tris-HCl pH8

0.06% w/v Bromophenol blue

Biotin dNTP mix (10x concentrate)

0.2mM dATP (Pharmacia)

0.2mM dCTP (Pharmacia)

0.2mM dGTP (Pharmacia)

0.1mM dTTP (Pharmacia)

0.1mM biotin-16-dUTP (Boehringer)

Stored at –20°C

Cresol Red

Prepared as a stock solution of 0.845mg/ml of the sodium salt in 1mM Tris-HCL, 0.1mM Na_2EDTA, pH8.8

Stored at 4°C

Denaturing solution for microsatellite screening

1.5M NaCl

0.5M NaOH

Prepared fresh as required

dNTPs

Prepared as a stock solution from Pharmacia Ultrapure dNTP set, containing either 1.5mM each, or 5mM each of dATP, dCTP, dGTP and dTTP

Stored at –20°C

Digoxigenin dNTP mix (10x concentrate)

0.2mM dATP (Pharmacia)

0.2mM dCTP (Pharmacia)

0.2mM dGTP (Pharmacia)

0.1mM dTTP (Pharmacia)

0.1mM digoxigenin-11-dUTP (Boehringer)

Stored at –20°C

FISH blocking wash solution

4x SSC

5% w/v skimmed milk powder

0.05% v/v Triton X-100

The solution of SSC and milk powder is filtered through Whatman no. 4 filter paper before the Triton is added and is prepared fresh for each use.

FISH Detection antibodies

Prepared fresh for each use and kept out of direct light

Detection layer 1:

6µl Texas red avidin DCS (4mg/ml, Vector)

1.5ml Blocking wash solution

Detection layer 2:

12.5µl biotinylated anti-avidin D (goat, 4mg/ml, Vector)

3µl FITC anti-digoxigenin (mouse, 4mg/ml, Vector)

1.5ml Blocking wash solution

Detection layer 3:

6µl Texas red avidin DCS (4mg/ml, Vector)

6µl FITC anti-mouse (goat, 4mg/ml, Vector)

1.5ml Blocking wash solution

FISH Hybridisation buffer

4x SSC

0.2% v/v Tween 20

20% w/v Dextran sulphate

Stored at 4°C

FISH nick translation buffer, 10x concentrate

500mM Tris-HCl pH7.8

50mM $MgCl_2$

100mM β-mercaptoethanol

100µg/ml nuclease-free BSA

Stored at –20°C

3:1 Fixative

150ml Methanol (AnalaR grade, BDH)

50ml Glacial acetic acid (AnalaR grade, BDH)

Prepared fresh for each use, kept on ice and discarded after 2 hours

LB agar (Luria agar)

10g/l Bactotryptone (Difco)

5g/l Yeast extract

10g/l sodium chloride

15g/l agar

pH7.5

Sterilised by autoclaving

LB broth (Luria broth)

10g/l Bactotryptone (Difco)

5g Yeast extract

10g sodium chloride

pH7.5

Sterilised by autoclaving

Modified Church's buffer

7% w/v sodium dodecyl sulphate

500mM NaH_2PO_4

1mM EDTA

0.1% w/v bovine serum albumin

Adjust to pH7.2 with glacial acetic acid

Stored at –20°C

Neutralising solution for microsatellite screening

1.5M NaCl

0.5M Tris-HCl, pH7.4

Prepared fresh as required

PCR Buffer 1, 10x concentrate (Perkin Elmer GeneAmp)

100mMTris-HCl, pH8.3,

500mM KCl,

15mM $MgCl_2$,

0.01% (w/v) gelatine

Stored at –20°C

PCR Buffer 2, 10x concentrate (Perkin Elmer GeneAmp)

100mMTris-HCl, pH8.3

500mM KCl

Stored at –20°C

Polyamine buffer, modified

80mM potassium chloride (KCl)

20mM sodium chloride (NaCl)

7.5mM Tris-base

2mM EDTA

0.5mM EGTA

0.2mM spermine tetrahydrochloride

0.5mM spermidine trihydrochloride

3mM dithiothreitol (DTT)

0.25% Triton X-100 (v/v)

pH7.2

Prepared fresh as required

Polyamine potassium chloride

75mM potassium chloride

0.2mM spermine tetrahydrochloride

0.5mM spermidine trihydrochloride

Stored at 4°C

Primers - shown 5' - 3'

Sequencing primers

T3 AATTAACCCTCACTAAAGGG

T7 GTAATACGACTCACTATAGGGC

Degenerate repeat primers

DRP1 GTGTGTGTGTGTGTGTGTGTA

DRP2 GTGTGTGTGTGTGTGTGTGTC

DRP3 GTGTGTGTGTGTGTGTGTGTT

DRP4 TGTGTGTGTGTGTGTGTGTGA

DRP5 TGTGTGTGTGTGTGTGTGTGC

DRP6 TGTGTGTGTGTGTGTGTGTGG

Reducing solution for Sanger PCR protocol

10% v/v β-Mercaptoethanol (Biorad) in Dilution buffer.

Stored on ice; use within 3 hours

Sanger dilution buffer

0.1mM Tris-base pH8.8

0.01mM Na_2EDTA

4.5μg/ml Cresol Red

2mM NaOH

Stored at 4°C

Sanger PCR buffer (10x concentrate)

450mM Tris-HCl, pH8.8

0.05mM Na_2EDTA

15mM $MgCl_2$

100mM $(NH_4)_2SO_4$

0.42mg/ml Cresol Red, (Sodium salt)

Stored at 4°C

Sequencing gel loading buffer

50μl deionised Formamide (Fluka)

10μl Blue Dextran solution (Perkin Elmer) (50mM EDTA, 50mg/ml Blue Dextran)

Stored at –20°C

SSC, 20x concentrate

3M sodium chloride

0.3M sodium citrate

Sterilised by autoclaving

TAE, 50x concentrate

40mM Tris base

5.71% v/v glacial acetic acid

2mM Na_2EDTA

Sterilised by autoclaving

TBE, 10x concentrate

89mM Tris base

89mM boric acid

2.5mM EDTA

Appendix 2

The following tables give the primer sequences and PCR conditions for each of the markers used in this study.

Table i Primers used in chromosome X mapping

Table ii Primers used in chromosome 1 mapping

F8c and F9 – E. Ostrander, personal communication

PGK1 and CHM – Deschênes *et al.* (1994)

PDHA1 - TOAST primer pair 2, Jiang *et al.* (1998)

1E7 - N. Suter, thesis

H62 and K338 - N. Holmes, personal communication

Key:

cpd = compound repeat

nd = not determined

td = touchdown

td59 $^{ext\ 2'}$ = touchdown 59°C, with a 2 minute extension for each cycle.

Table I

Marker name	Cytogenetic location	Feature amplified	Primers:		bp	t°C / t_{ann}	mM [Mg^{++}]
x1	Xq24 - 25	$(CA)_{11}$	x1tggf	AGTGTAGGTGGAGGATTAGGG	342 - 346	55	1
			x1tgg	GAGTATTTCAAGCACACAGCCC			
x9	Xp13	$(AC)_{16}$	x9tgaf	TTGAGAGAAGGGAGTGTATTTGCT	224 - 228.	57 / 50	1.5
			x9tgar	ACTGGAGGTTTCGTTTATGGATAT			
x12	nd	$(CT)_4GT(CT)_7$	x12t3f	CTCCCTCTTCCTCTGCCC	138 - 142	57 / 55	1.5
			x12t3r	GAGCTCCAAAAGGAATGCA			
x20	Xp21 & Xq21	$(GT)_{17}$	x20t3f	AGAAAAAGGTACACCATGCCGC	326 - 344	60	1.5
			x20t3r	ACAAAAAGACGATGCTGATGGG			
x21	Xp24	$(GA)_{12}$	x21t7f(2)	AAAAATATACATACAAACACAC	99 - 110	td46 / 55	1.5
			x21t7r(2)	CTGAGAAAATACACACAAGAT			
x24	Xq12	$(TG)_{20}$	x24caaf	AGGATTTCCATCTGTCTCATAGC	143 - 158	57	1.5
			x24caar	ATTGAAGCATTATTCACAATAGCC			
x27	Xq13 - 21	$(AC)_{16}A(AC)_4$	x27f	GGGAGGTCGTGCAATCAG	146 - 166	57 / 60	2 / 1.5
			x27r	GGAGCAGGCAGAGAAGATGA			
x28	Xp23 - 24	$(CTTT)_{20}$	x28f	GCTGACTGCCTGCAAAAGGC	222 - 262	60	1.5
			x28r	GTGATGTAGCCAGAAAACAGCG			
x30	Xp12 - 13	$(AC)_{11}AT(AC)_6$	x30gtt	TGAGCGCGGTGACATCTCAG	184	td49	1.5
			x30gttr	CCACACCTGCTCCAACAC			
x31	Xp23	$(GT)_{14}$	x31t7f	GGATCAAGTCCTGCATCGGG	355	td60	1.5
			x31t7r	GAGGAAAAGGAGGTGCTGTCCG			
x33	Xq21.2 - 22	$(AC)_{12}$	x33gtaf	TTGTGAAGAAGGCAAAGTTGG	298 - 300	55	1.5
			x33gtar	TGAAAGAGGAAAGCCACTGA			
x34	Xp11 - 12	$(AC)_8$ cpd	x34tgaf	TCCATCTCAGAGTCAGTTGCC	189	55	1
			x34tgar2	GATTCCCAAGTTCAATACCCC			
x35	Xq12	$(CA)_{17}$	x35gtaf	AATGGAAGGGCACTTGAAAA	151	55 / 60	1
			x35gtar2	AAATGATGGGTCTTTGTGTGTG			
x40	Xq21.3 - 23	$(CAAT)_7$	x40f	GTAGCCCCATTTGTTTATATTTGC	190	60 / 60	1.5
			x40r	GATCACTGTCTTACACCACAGGC			
x48	nd	$(GT)_{17}$	x48actf	ATAGTTCATTCTTGTTTTTACTCC	95	td57	1.5
			x48actr	TTTCAGAGTATTAGAACCAATGCC			
x49	Xq11	$(AC)_{12}$	x49tga	AAACAACAACAGCAGATACCACA	190	57	1.5
			x49tgar	TTTGTCCTTTTGCTTTCTGAGT			
x50	Xq22 -23	$(GT)_2CT(GT)_2AT(GT)_8$	x50cagf	ATGACAACAGCATCACAGTCCC	93 - 95	td57 / 55	2.5 /1.5
			x50cagr	GCAAGTTGGTTAGCATGTCAGG			
x51	Xp21 -23	$(AC)_{18}$	x51gtcf	CATACTCCCAGCCAAGTGTCAG	100 - 122	60 / 60	1.5
			x51gtcr	TCCAGGTGGAGAGCAACTGG			
x53	nd	$(AC)_{19}$	x53gtaf	CCAGGTACTGAGTGATATAGT	302 - 312	53	1.5

Table I

Marker name	Cytogenetic location	Feature amplified	Primers:		bp	$t^{o}C$ / t_{ann}	mM [Mg^{++}]
x55	nd	$(AC)_{13}$	x53gtar	TAATAGTGTCCGCGTTAGGAA	218 - 230	57	1.5
			x55t7f	GCAAATGACATATCCTATAAGGGG			
			x55t7r	AGGATCGAGTCCCAAGTCGG			
x56	Xq21.2 - 21.3	$(GA)_{11}$	x56t7f	CTCCTTCCCAAGGTTTTGCC	120 - 122	57	1
			x56t7r	CTCCCTCTGCTATCCTCCCC			
x57	Xp11 - 12	$(AC)_{14}$	x57gta	TAGATTGGTAGCTTCAGAGGGG	222	57	1
			x57gtar	GCAAGTCAAAGGTAGGGCTG			
x58	Xq26 - 28.1	$(AC)_{16}$	x58t7f	GTTAGATGTGTTGGAAGCAGGG	385 - 378	57	1
			x58t7r	AGCAAGCAAAGAAAGTGGGG			
x64	Xq11	$GA_2(CA)_{13}(GA)_{20}$	x64gtc	CCATATCTTGTAAGCAGGAATCA	321 - 328	57	1.5
			x64gtcr	TTAGCACCACCTTCCACCC			
x66	Xq21.2 - 21.3	$(GT)_{16}$	x66gttf	AAGCACCTATTACAATCCATACCG	257 - 265	54	1.5
			x66acgf	GCTTTGCCTTTGACAGTTATCG			
x67	Xp13 - 21	$(AC)_4AT(AC)_{12}$	x67gta	AAGGGATGCTTCTCCAGATAAA	230 - 242	td57 / 60	1
			x67gtar	TGAGTAGACAAGGCAAACACCC			
x68	Xq26 - 27	$(CA)_{14}$	x68f	AATGGATTTCATGATTATTGGCC	315 - 326	57	1.5
			x68r	GGAGACTGTTTTTGTGGCCC			
x69	Xq21.2 - 22	$(CA)_{17}$	x69f	TGAAGGGTAGGGAAGATAGTGG	84	53 / 60	1.5
			x69gtar	GGATGGATGAAGATGTGGTA			
x71	Xp21	$(AC)_{15}$	x71f	GCAAATTGTGGACATAGCCC	152	57	1.5
			x71r	ATCTAGTTCCATCTGTGTTGTTGC			
x74	Xp21	$(TG)_{20}$	x74gttf	ATGGCCCGTCATAGGTAGGT	213	57	1.5
			x74gttr	TCCAGCCATTGAAAAAGAGAAG			
x77	Xq13	$(CA)_8C(AC)_5$	x77tgcf	GCTGATGAATGAATAGAGAGG	163	50	1
			x77tgcr2	AGCACAGAGATGGACACGG			
x78	Xp23 - 24	$(AC)_{11}T(GA)_{14}$	x78f	CAACTGAGCTATGCAGGCG	187	td57 / 60	1.5
			x78r	GATCGAGTCCCACATCGG			
x91	Xp21 - 22	AC_{12} & TA_{20+}	x91gtar	CAGCATAAAAAAAAAAATCACTCA	195 - 244	55	1.5
			x91gtaf	TAAAGCCAGATAACTTGAGAAAA			
CHM*	Xq13 - 21*	$(TG)_{17}(AG)_6$	CHM-TG	GGGAAGGTGGCAGTAGCAAGG	170 - 172	55	1.5
			CHM-AC	CAGAGACTAATGAGGGGGCTCC			
PGK1*	Xq13 - 21*	$(TG)_{11}CA(TG)_2CA(TG)_6$	PGK-TG	AATATTTTTCATGTATAGCACTCC	181 - 183	td57	1.5
			PGK-AC	ATGCTCCCTTCAGTTCTTTCC			
PDHA1[@]	ND	STS	PDHA1F2	TTACAACATGGCAGCTTTGTG	386	td57	1.5
			PDHA1R2	CTTCCCAGATCTACAGTAGGCAG			
F8c[+]	ND	(GAAA)n - CPD	F8c -1	GATCTAGTGCCTGTAAAAAC	267 - 329	td59 [and 2']	1.5
			F8c -3	CCTGGCTCCCTAGAACCACG			

Table I

Marker name	Cytogenetic location	Feature amplified	Primers:		bp	$t^{\circ}C$ / t_{anm}	mM [Mg^{++}]
$F9^{+}$	ND	$(TTTA)_n$	F9-3	GAACAGATTAGTAGCTTGAATTGGAC	138 - 158	60	1.5
			F9-4	AGAATTCAGACTGAGAACAATTACAG			
$1E7^{++}$	Xp24++	$(TG)_{21}$	1E7F	CGTGCCCTCTGCAAAGTGC	85	60	1.5
			1E7R	CTGACCGCATGACTTATCC			
$H62^{a}$	Xq22a	$(CA)_{13}$	H62F	TTCTCTCCCTAGAAGCAGTTGG	266 - 278	td53	1.5
			H62R	TAGTGCTGGTTGGTGTATCATTG			

Table ii

Marker name	Cytogenetic location	Feature amplified	Primers:		bp	t°C	mM [Mg^{2+}]
1-01	1q13.1 - 14.1	Anon. STS	1-01t3for	AAACATTCTTCCACTACAACGC	136	57	1.5
			1-01t3rev	TAGGAATTCAAACGGAACGG			
1-03	1q14	Anon. STS	1-03t3for	AATATTACAGTACTTGGGGCTTCA	110	td55	1.5
			1-03t3rev	GAATCCTGGACAAACCTTCATC			
1-09	1q32	Anon. STS	1-09t3for	GTGGGGAGAGTGGGACATC	111	td57	1.5
			1-09t3rev	TACAAGCCATTTATGGAAAGCC			
1-11	1q23 - 24	Anon. STS	1-11t3f	TGTTGGAAAAATTGGTATTGTTGG	103	57	1.5
			1-11t3r	GGCTGGCTGATGGAAAATAATG			
1-20	1q21 - 23	Anon. STS	1-20t3for	CCAGGCATTACTGTTTTCAGC	209	57	1.5
			1-20t3r2	TCTTTCTTTCCCCTCATCACA			
1-21	1q35	Anon. STS	1-21t3for	AAATAAATACGGATGCCTGGG	156	60	1.5
			1-21t3r	AGAGACACACGGAGAGAGGC			
1-26	nd	(GT)13 cpd	1-26for	AGTAAGAAATGTGTGCATGTGCG	217	td57	1.5
			1-26rev	GTCCGGTCCACCATAGTCCC			
1-27	1q35.2 - 35.3	Anon. STS	1-27t3for	TCAGGGAGCTCAGGACGC	142	td60	1.5
			1-27t3rev	TCAGGATGCACTACTTCAGGG			
1-28	1q31.1 - 31.2	$(GT)_{17}$	1-28t7f	ACTGAAAGATCAGGTGTTAGGTAGC	144	td57	1.5
			1-28t7r	TATCCCTTGGCCCCTCCT			
1-30	1q14.3 - 21	Anon. STS	1-30t3for	ATTAGTTGCTCCCACAGTAAAAGG	109	58	1.5
			1-30t3rev	GTTTTAGGGACCCAGAGTGTTG			
1-32	1q34 - 35.2	Anon. STS	1-32t7f	TGTACTCTCGGGCAGGGG	246	td57	1.5
			1-32t7r	AAAGGTGCCATTTTATTATTTATCC			
1-33	1q31.2 - 31.3	$(CA)_{18}$	1-33for	GAAGTGAACATAGGATTTTTCTGA	233	57	1.5
			1-33rev	AGATGTTAATAGCTGGAGATTGTG			
1-34	1q31.1 -31.3	Anon. STS	1-34t3for	TTTGCTGAAAAAAACAAAGGG	319	td61	1.5
			1-34t3rev	CCTGTCAATTCCAGGGTAGC			
1-42	1q32 - 33	Anon. STS	1-42t7for	TCTCCCTGTCCCACTGCC	114	td60	1.5
			1-42t7rev	CGCTTGAACCACTGAGCC			
1-43	1q22-23	Anon. STS	1-43t3for	CCAACCCAGAACAACAGAGAC	190	60	1.5
			1-43t3r2	GAGCTGCAGGCAGGAAGAG			

Table ii

Marker name	Cytogenetic location	Feature amplified	Primers:		bp	t°C	mM [Mg^{2+}]
1-45	1q33 - 34	Anon. STS	1-45t3for	CAAAAAATGGGGATTTTCACCT	171	60	1.5
			1-45t3r2	GAGCTGCAGGCAGGAAGAG			
1-46	1q12	Anon. STS	1-46t3f	TGTTTGGAGGGCATCAGC	138	td57	1.5
			1-46t3r	AAGCCAGGACTTGATTTATGATTG			
1-49	1q31	Anon. STS	1-49t3for	TGCATGTTTCAATCAATGTGG	107	td57	1.5
			1-49t3rev	AGACAGCTTTCCTGGAGTTTG			
1-57	1q31.2 - 31.3	t3 - $(TTTC)_{15}(TTTTC)_4(TTTC)_2$	1-57t3for	GCTAAGGTCGATGATGGGTT	247	td57	1.5
			1-57t3rev	CTTGGTTGCCAGTACCTTTG			
1-58	1q33 - 34	Anon. STS	1-58t3for	TTAATGGAGGTCATCACAGCA	201	60	1.5
			1-58t3rev	TTCACTTTGGGAGGGATCAG			
1-60	1q12	Anon. STS	1-60t3for	CCCAACCTCAATTCCATTCT	105	td57	1.5
			1-60t3rev	GTGGAAACACTATGCCCAGC			
H32	1q32 - 34	Anon. STS	H32T3F	TCTGGCTTTATCAGGAGTTTTTCT	150	60	1.5
			H32T3R	GGTAATGAAGTGACTATTTCCCCA			
H38	1q12	Anon. STS	H38T3F	GCTCAGTTTAAAGCAATATGTCCA	112	60	1.5
			H38T3R	TGACTCTCCTTCCCACGC			
H152	1q34 - 35.1	Anon. STS	H152T3F	TTTTCTGGGCTGCAAAATG	102	60	1.5
			H152T3R	GTTTCCACGGATGCTCTTATTC			
H233	1q34	Anon. STS	H233T3F	AAAGTTTCAAAAAACCCTGTCCTC	145	60	1.5
			H233T3R	TGTCACAAAAAATAACTAGGTTCAGC			
H254	1q35.1 - 35.2	Anon. STS	H254T3F	AGTCTTCCAAACAGTGAAAAACAA	103	60	1.5
			H254T3R	AAGGCAAATCATTCCACCAG			
K338[a]	1q11 - 12	$(GT)_{12}$	K338f	ACAAAGCAATTTTTGCTACAAGC	(250)	57	1.5
			K338r	TAATAGGTGTGGGACATTCGTT			

Appendix 3

Examples of BLAST outputs from searches of the GenBank database

The following are six examples of the results of searching the GenBank database using the BLAST programme. In each case, a maximum of the ten most significant matches are shown.

1. The first example is of sequence from x24. This had significant similarity to rabbit, human and mouse sequences from the phosphorylase kinase, alpha 1 gene (PHKA1) located on chromosome X, with the most significant matches being from the GenBank entry to rabbit sequences PHKA1 mRNA sequences, accession numbers J023247 and M64656,
2. The following example is of sequence from x57, which showed significant similarity to human (accession numbers AF000993, AF000992) and mouse (accession numbers AJ002730, AC006508) sequences for ubiquitously transcribed tetratricopeptide repeat gene on chromosomes X and Y (UTX). This sequence therefore not only shows homology with a known gene, but the gene is known to be located on chromosomes X and Y, providing additional evidence that this is derived from chromosome X.
3. The sequence of clone x74 showed significant homology to sequence from the human dystrophin gene, (DMD, accession number U60822). This homology was confirmed experimentally when the x74 clone was used in FISH experiments on metaphase spreads from a female dog known to be a carrier of this disease gene. The FISH analysis confirmed that this animal had one normal X, in which both chromatids contained the DMD sequence, and one chromosome X where the DMD region was deleted. PCR analysis using primers from the x74 clone provided additional confirmation, when they failed to amplify a product from DNA from the male offspring (i.e. affected) of the carrier bitch (data not shown).
4. The sequence of x92 showed some similarity to human (accession number AA282065) and mouse (accession number U43892) sequences from the ATP-binding cassette 7-iron transporter gene (ABC7), located on chromosome X.
5. The next example shows a sequence with significant homology to a known chromosome X-derived sequence, (accession number U69729). The query sequence is from clone x13 and it has a significant database match to a human Xp22-derived cosmid sequence (accession number: U69729). This gives additional confirmation that x13 is derived from chromosome X.
6. Finally an example of a search result from sequence containing a repetitive element, in this case, sequence from x56 that contains a SINE element.

All the examples shown are from the small-insert chromosome X-enriched library since there were no significant matches from the chromosome 1-enriched library except for repetitive elements, which were similar to those found in the chromosome X library.

1. Query= x24
(370 letters)
Score E
Sequences producing significant alignments: (bits) Value

GB:RABPKA	J03247	Rabbit phosphorylase kinase (alpha s...	475	2.1e-29
GB:RABPLASISM	M64656	Rabbit phosphorylase kinase alpha su...	475	2.1e-29
GB:HSPHKA1	X73874	H.sapiens PHKA 1 mRNA.	466	1.2e-28
GB:MMPHOKA1	X74616	M.musculus mRNA for phosphorylase ki...	433	7.9e-26
GB:MMPHOKA1	X74616	M.musculus mRNA for phosphorylase ki...	433	7.9e-26
GB:HSB50D112	F00757	H. sapiens partial cDNA sequence; cl...	427	8.7e-26
GB:HUMPHKA2	D38616	Human mRNA for phosphorylase kinase ...	271	4.0e-12
GB:HSPHKLA	X00497	H.sapiens PHKLA mRNA.	271	4.0e-12
GB:H31848	H31848	EST106346 Rattus sp. cDNA 5' end.	139	0.30
GB:MMPHKA1I	X73877	M.musculus Phka1 unprocessed pre-mRN...	133	0.73

>GB:RABPKA J03247 Rabbit phosphorylase kinase (alpha subunit) mRNA, complete cds.
Length = 4441

Minus Strand HSPs:

Score = 475 (131.3 bits), Expect = 2.1e-29, P = 2.1e-29
Identities = 99/104 (95%), Positives = 99/104 (95%), Strand = Minus / Plus

```
Query:   237 CATAGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCC 178
             || |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Sbjct:  1002 CAAAGGCAGCCCTGGAAGCATTAGATGAACTTGACCTGTTTGGTGTGAAAGGTGGGCCCC 1061

Query:   177 AATCAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             ||||||||||||||||| | ||||||||||||||||||||||||
Sbjct:  1062 AATCAGTTATCCATGTCTTGGCTGATGAAGTACAACACTGCCAG 1105
```

>GB:RABPLASISM M64656 Rabbit phosphorylase kinase alpha subunit mRNA, complete cds.
Length = 4458

Minus Strand HSPs:

Score = 475 (131.3 bits), Expect = 2.1e-29, P = 2.1e-29
Identities = 99/104 (95%), Positives = 99/104 (95%), Strand = Minus / Plus

```
Query:   237 CATAGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCC 178
             || |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Sbjct:  1056 CAAAGGCAGCCCTGGAAGCATTAGATGAACTTGACCTGTTTGGTGTGAAAGGTGGGCCCC 1115

Query:   177 AATCAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             ||||||||||||||||| | ||||||||||||||||||||||||
Sbjct:  1116 AATCAGTTATCCATGTCTTGGCTGATGAAGTACAACACTGCCAG 1159
```

>GB:HSPHKA1 X73874 H.sapiens PHKA 1 mRNA.
Length = 4215

Minus Strand HSPs:

Score = 466 (128.8 bits), Expect = 1.2e-28, P = 1.2e-28
Identities = 98/104 (94%), Positives = 98/104 (94%), Strand = Minus / Plus

```
Query:   237 CATAGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCC 178
             || |||||||||||||||||||| |||||||||| |||||||||||||||||||||||| |
Sbjct:   775 CAAAGGCAGCCCTGGAAGCATTAGATGAACTGGATCTGTTTGGTGTGAAAGGTGGGCCTC 834

Query:   177 AATCAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             ||||||||||||||||||| |||||||||||||| |||||||||
Sbjct:   835 AATCAGTTATCCATGTCCTGGCTGATGAAGTACAGCACTGCCAG 878
```

>GB:MMPHOKA1 X74616 M.musculus mRNA for phosphorylase kinase.
Length = 4130

Minus Strand HSPs:

Score = 433 (119.6 bits), Expect = 7.9e-26, P = 7.9e-26
Identities = 93/101 (92%), Positives = 93/101 (92%), Strand = Minus / Plus

```
Query:   234 AGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCCAAT 175
             |||||||||||||||| | |||||| | ||| ||||||||||||||||||||| ||||
Sbjct:   835 AGGCAGCCCTGGAAGCACTAGATGAATTAGACTTGTTTGGTGTGAAAGGTGGGCCACAAT 894
```

```
Query:   174 CAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             ||||||||||||||||| |||||||||| ||||||||||
Sbjct:   895 CAGTTATCCATGTCCTGGCTGATGAAGTCCAACACTGCCAG 935


>GB:MMPHOKA1 X74616 M.musculus mRNA for phosphorylase kinase.
Length = 4130

Minus Strand HSPs:

Score = 433 (119.6 bits), Expect = 7.9e-26, P = 7.9e-26
Identities = 93/101 (92%), Positives = 93/101 (92%), Strand = Minus / Plus

Query:   234 AGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCCAAT 175
             ||||||||||||||||| | |||||| | ||| ||||||||||||||||||||| ||||
Sbjct:   835 AGGCAGCCCTGGAAGCACTAGATGAATTAGACTTGTTTGGTGTGAAAGGTGGGCCACAAT 894

Query:   174 CAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             ||||||||||||||||| |||||||||| ||||||||||
Sbjct:   895 CAGTTATCCATGTCCTGGCTGATGAAGTCCAACACTGCCAG 935


>GB:HSB50D112 F00757 H. sapiens partial cDNA sequence; clone 50D11; version 1;
strand(+), single read.
Length = 338

Minus Strand HSPs:

Score = 427 (118.0 bits), Expect = 8.7e-26, P = 8.7e-26
Identities = 91/99 (91%), Positives = 91/99 (91%), Strand = Minus / Plus

Query:   232 GCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCCAATCA 173
             |||||||||||||||| |||||||||  || |||||||||||||| ||||| ||||||
Sbjct:     1 GCAGCCCTGGAAGCATTAGATGAACTGGTTCTNTTTGGTGTGAAAGGNGGGCCTCAATCA 60

Query:   172 GTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             ||||||||||||||| ||||||||||||||| |||||||||
Sbjct:    61 GTTATCCATGTCCTGGCTGATGAAGTACAGCACTGCCAG 99


>GB:HUMPHKA2 D38616 Human mRNA for phosphorylase kinase alpha subunit, complete
cds.
Length = 4392

Minus Strand HSPs:

Score = 271 (74.9 bits), Expect = 4.0e-12, P = 4.0e-12
Identities = 75/101 (74%), Positives = 75/101 (74%), Strand = Minus / Plus

Query:   234 AGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCCAAT 175
             ||||||| || || ||| | |||||||||||||| ||||| |   | || || | | | |
Sbjct:   788 AGGCAGCTCTTGAGGCAATTGATGAACTGGACCTTTTTGGAGCCCATGGAGGACGCAAGT 847

Query:   174 CAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             |||| || ||||| ||  | ||||| ||  | |||||||||
Sbjct:   848 CAGTGATTCATGTTCTGCCAGATGAGGTCGAGCACTGCCAG 888


>GB:HSPHKLA X80497 H.sapiens PHKLA mRNA.
Length = 4566

Minus Strand HSPs:

Score = 271 (74.9 bits), Expect = 4.0e-12, P = 4.0e-12
Identities = 75/101 (74%), Positives = 75/101 (74%), Strand = Minus / Plus

Query:   234 AGGCAGCCCTGGAAGCATTGGATGAACTGGACCTGTTTGGTGTGAAAGGTGGGCCCCAAT 175
             ||||||| || || ||| | |||||||||||||| ||||| |   | || || | | | |
Sbjct:   743 AGGCAGCTCTTGAGGCAATTGATGAACTGGACCTTTTTGGAGCCCATGGAGGACGCAAGT 802

Query:   174 CAGTTATCCATGTCCTAGCTGATGAAGTACAACACTGCCAG 134
             |||| || ||||| ||  | ||||| ||  | |||||||||
Sbjct:   803 CAGTGATTCATGTTCTGCCAGATGAGGTCGAGCACTGCCAG 843
```

2. Query= **x57**
(612 letters)

Sequences producing significant alignments:	Score (bits)	E Value
EM:AF000993 AF000993 Homo sapiens ubiquitous TPR motif, X isofo...	230	5e-58
EM:AF000992 AF000992 Homo sapiens ubiquitous TPR motif, X isofo...	230	5e-58
EM:MMAJ2730 AJ002730 Mus musculus mRNA for ubiquitously transcr...	196	8e-48
EM:AF000994 AF000994 Homo sapiens ubiquitous TPR motif, Y isofo...	164	3e-38
EM:AC006508 AC006508 Mus musculus Yp BAC GSMB-187H15 (Genome Sy...	88	2e-15
EM:MMUTY Y09222 M.musculus Uty mRNA for male-specific histocomp...	78	2e-12
EM:AF057367 AF057367 Mus musculus male-specific histocompatibil...	78	2e-12
EM:CEUC46G7 U97593 Caenorhabditis elegans cosmid C46G7.	46	0.010
EM:CEC54C6 Z77131 Caenorhabditis elegans cosmid C54C6	40	0.65

>EM:AF000993 AF000993 Homo sapiens ubiquitous TPR motif, X isoform (UTX) mRNA, alternative transcript 2, complete cds.
Length = 5418

Score = 230 bits (115), Expect = 5e-58
Identities = 123/126 (97%)
Strand = Plus / Minus

```
Query: 48   tcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccatataatctct 107
            |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct: 4030 tcacaaatgctacagtaatgagctggttcttcttttgtccgcccatgccatataatctct 3971


Query: 108  tttcctgcagcaatgagagcttccctcaatgtctgacattgcttcagagttctcanaaga 167
            |||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||
Sbjct: 3970 tttcctgcagcaatgagagcttccctcaatgtctgacattgcttcagagttcttagaaga 3911


Query: 168  caatac 173
            ||||||
Sbjct: 3910 caatac 3905
```

>EM:AF000992 AF000992 Homo sapiens ubiquitous TPR motif, X isoform (UTX) mRNA, alternative transcript 1, complete cds.
Length = 4856

Score = 230 bits (115), Expect = 5e-58
Identities = 123/126 (97%)
Strand = Plus / Minus

```
Query: 48   tcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccatataatctct 107
            |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct: 4030 tcacaaatgctacagtaatgagctggttcttcttttgtccgcccatgccatataatctct 3971


Query: 108  tttcctgcagcaatgagagcttccctcaatgtctgacattgcttcagagttctcanaaga 167
            |||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||
Sbjct: 3970 tttcctgcagcaatgagagcttccctcaatgtctgacattgcttcagagttcttagaaga 3911


Query: 168  caatac 173
            ||||||
Sbjct: 3910 caatac 3905
```

>EM:MMAJ2730 AJ002730 Mus musculus mRNA for ubiquitously transcribed TPR gene on the X chromosome (Utx)
Length = 4215

Score = 196 bits (98), Expect = 8e-48
Identities = 121/129 (93%)
Strand = Plus / Minus

```
Query: 45   acctcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccatataatc 104
            |||||||||| ||||| ||||| ||||||||||||||||||||| |||||||| |||
Sbjct: 3805 acctcacaaatactacaataatgagctggttcttcttttgtccgcccgtgccatatgatc 3746


Query: 105  tcttttcctgcagcaatgagagcttccctcaatgtctgacattgcttcagagttctcana 164
            |||||||||||||||| |||||||||||||||||||||||||||||| ||||||| |
Sbjct: 3745 tcttttcctgcagcaattagagcttccctcaatgtctgacattgcttcagcgttctcaga 3686
```

```
Query: 165  agacaatac 173
            |||||||||
Sbjct: 3685 agacaatac 3677
```

>EM:AF000994 AF000994 Homo sapiens ubiquitous TPR motif, Y isoform (UTY) mRNA, alternative transcript 3, complete cds.
Length = 6476

Score = 164 bits (82), Expect = 3e-38
Identities = 117/129 (90%)
Strand = Plus / Minus

```
Query: 45   acctcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccatataatc 104
            ||||||||||||||||||||||| |||||||| || ||||||||||||||||||||||| |
Sbjct: 4853 acctcacaaatgctacagtaatgagctggttcatcatttgtccgcccatgccatataacc 4794

Query: 105  tcttttcctgcagcaatgagagcttccctcaatgtctgacattgcttcagagttctcana 164
            |||||||||| ||||  |||||||| |||||||||||| |||||||||| || ||| |
Sbjct: 4793 tcttttcctgctgcaacaagagcttctctcaatgtctgatattgcttcagaattttcaaa 4734

Query: 165  agacaatac 173
            |||||||||
Sbjct: 4733 agacaatac 4725
```

>EM:AC006508 AC006508 Mus musculus Yp BAC GSMB-187H15 (Genome Systems Mouse BAC Library) complete sequence.
Length = 185286

Score = 88.4 bits (44), Expect = 2e-15
Identities = 95/112 (84%)
Strand = Plus / Minus

```
Query: 40     tacttacctcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccata 99
               |||||||||||||| |||||||||||| | |||||||| || ||| ||| ||||||||| |
Sbjct: 184262 tacttacctcacagatgctacagtaaggagctggttcatcatttatcctcccatgccaca 184203

Query: 100    taatctcttttcctgcagcaatgagagcttccctcaatgtctgacattgctt 151
               || |||||| ||||| |||| ||||||||| ||||| |||||||| |||||
Sbjct: 184202 aaacctctttccctgctgcaacaagagcttctctcaaagtctgacaatgctt 184151
```

>EM:MMUTY Y09222 M.musculus Uty mRNA for male-specific histocompatibility antigen H-YDb
Length = 3823

Score = 78.4 bits (39), Expect = 2e-12
Identities = 90/107 (84%)
Strand = Plus / Minus

```
Query: 45   acctcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccatataatc 104
            |||||||| |||||||||||| | |||||||| || ||| ||| ||||||||| | || |
Sbjct: 3511 acctcacagatgctacagtaaggagctggttcatcatttatcctcccatgccacaaaacc 3452

Query: 105  tcttttcctgcagcaatgagagcttccctcaatgtctgacattgctt 151
            ||||| ||||| |||| |||||||| ||||| |||||||| |||||
Sbjct: 3451 tctttccctgctgcaacaagagcttctctcaaagtctgacaatgctt 3405
```

>EM:AF057367 AF057367 Mus musculus male-specific histocompatibility antigen H-YDb (Uty) mRNA, complete cds.
Length = 5219

Score = 78.4 bits (39), Expect = 2e-12
Identities = 90/107 (84%)
Strand = Plus / Minus

```
Query: 45   acctcacaaatgctacagtaatgcgctggttcttcttttgtccgcccatgccatataatc 104
            |||||||| |||||||||||| | |||||||| || ||| ||| ||||||||| | || |
```

```
Sbjct: 3609 acctcacagatgctacagtaaggagctggttcatcatttatcctcccatgccacaaaacc 3550


Query: 105  tcttttcctgcagcaatgagagcttccctcaatgtctgacattgctt 151
             ||||| ||||| ||||  |||||||| ||||| |||||||| |||||
Sbjct: 3549 tctttccctgctgcaacaagagcttctctcaaagtctgacaatgctt 3503
```

3. Query= **x74**

(558 letters)

```
Score     E
Sequences producing significant alignments:                    (bits)  Value

GB:HSU60822   U60822 Human dystrophin (DMD) gene, exons 7...     372  5.6e-23
GB:HUMNRF1A   L24123 Homo sapiens NRF1 protein (NRF1) mRNA.      157  0.025
GB:I08319     I08319 Sequence 1 from Patent WO 8401173.         150  0.086
GB:CEC14B4    Z81471 Caenorhabditis elegans cosmid C14B4.       130  0.13
GB:HSHBZ17    X77366 H.sapiens HBZ17 mRNA.                     140  0.49
GB:HSTCF11A   X84060 H.sapiens TCF11 gene, exon 3-6.           140  0.49
GB:CHATTPROG  X68046 A.thaliana genes for chloroplast rib...   140  0.50
GB:SCYBR188C  Z36057 S.cerevisiae chromosome II reading f...   139  0.55
GB:YSCSUP46A  M88650 Saccharomyces cerevisia (clone pYsup...   139  0.55
GB:SCYBR189W  Z36058 S.cerevisiae chromosome II reading f...   139  0.55


>GB:HSU60822 U60822 Human dystrophin (DMD) gene, exons 7, 8 and 9, and partial
cds.
Length = 112,359

Minus Strand HSPs:

Score = 372 (102.8 bits), Expect = 5.6e-23, Sum P(3) = 5.6e-23
Identities = 132/204 (64%), Positives = 132/204 (64%), Strand = Minus / Plus

Query:    294 ATATCTCAATTTAGAGAACTGCTCTTATAATTGTAATTTTTTAGCAGAATCCTACCTATG 235
              |||||| |  ||| | | |   | ||| |||| | ||| ||| | ||||||||||| |
Sbjct:  99104 ATATCTGAGCTTAAAAAGCATTTTTTACAATTATCATTATTTGGTAGAATCCTACCCACC 99163

Query:    234 ACTGGGCCATAGTTTTAGTGAGATGGCAAGGCTTTCATTTGCTTGACTCTTCCAGCCCAC 175
              |||   |||||||||||||||||| |||||  ||||    |||||  |||| ||  || |
Sbjct:  99164 GCTGACTCATAGTTTTAGTGAGATGACAAGGAATTCACACACTTGAATCTTTCAAACCCC 99223

Query:    174 CACGGATTCAAACACGGTACAAAATTACTAACAAATGTGCTCCATCCAAGTTCTAAGGGG 115
              || | || || || |    || ||    |    ||  |   |   | ||||||||||| |
Sbjct:  99224 CAGGTATGCAGACGCTTAGCACAAAATTTTTGGAAGCTATGCTTCCAAAGTTCTAAGGAG 99283

Query:    114 TTATTTTTTTCTTTCCTGGGTCAG 91
              |    |||| || ||||  ||| |
Sbjct:  99284 TCTGCTTTTCCTATCCTCAGTCTG 99307

Score = 128 (35.4 bits), Expect = 5.6e-23, Sum P(3) = 5.6e-23
Identities = 42/63 (66%), Positives = 42/63 (66%), Strand = Minus / Plus

Query:     63 TCAATAATTTTTGAACATTGATTNCTATGTAAAGCACTGTGCTATTTTGTGGAATCTGTA 4
              |||||| | |||    |   |    | |||||||||| |||||||||||| |  | | |
Sbjct:  99346 TCAATATTATTTTTGAAAATACGTTTTTGTAAAGCACGGTGCTATTTTGTAGTGTGTATG 99405

Query:      3 CCA 1
              |||
Sbjct:  99406 CCA 99408

Score = 122 (33.7 bits), Expect = 0.0081, Sum P(3) = 0.0081
Identities = 38/55 (69%), Positives = 38/55 (69%), Strand = Minus / Plus

Query:    114 TTATTTTTTTCTTTCCTGGGTCAGTACTTGAACAATTTTTTTCTGTAAATATCAA 60
              | |||||||||| ||||||    ||| |||| | |   | | |  | |||||||
Sbjct:  28625 TAATTTTTTTCTTACCTGGGAGCCTACATGAATACTGAGTGTATTAAGATATCAA 28679

Score = 95 (26.3 bits), Expect = 3.0e-20, Sum P(3) = 3.0e-20
Identities = 23/28 (82%), Positives = 23/28 (82%), Strand = Minus / Plus

Query:     79 TTTTTTTCTGTAAATATCAATAATTTTT 52
              |||||||||| |||||| | | | ||
Sbjct:  99326 TTTTTTTCTGTAGATATCAGTCAATATT 99353

Score = 94 (26.0 bits), Expect = 5.6e-23, Sum P(3) = 5.6e-23
Identities = 22/26 (84%), Positives = 22/26 (84%), Strand = Minus / Plus

Query:    347 TCTATCATGTTAATTTTTCTTGGTTT 322
              |||||| |||||||||| |||  ||
Sbjct:  27136 TCTATCCTGTTAATTTTTATTGTATT 27161

Score = 80 (22.1 bits), Expect = 8.1e-22, Sum P(3) = 8.1e-22
Identities = 24/35 (68%), Positives = 24/35 (68%), Strand = Minus / Plus
```

```
Query:    495 TNTTTTTTTTTAACCCATTTTTTNGATATTACTAA 461
              | ||||||||||||    || |||  || | | |||
Sbjct:  56084 TTTTTTTTTTTAATGGTTTATTTCTATTTGAGTAA 56118
```

4. Query= **x92**
(456 letters)

```
Score     E
Sequences producing significant alignments:                    (bits)  Value

GB:AA282065 AA282065 zt02e02.s1 Soares NbHTGBC Homo sapie...    371  1.2e-25
GB:MMU43892 U43892 Mus musculus ABC transporter-7 mRNA, p...    335  2.2e-17
GB:MMU43892 U43892 Mus musculus ABC transporter-7 mRNA, p...    335  2.2e-17
GB:N79790   N79790 za95h03.s1 Homo sapiens cDNA clone 300...    140  0.37
GB:CPU30821 U30821 Cyanophora paradoxa cyanelle, complete...    117  0.77
GB:W07545   W07545 za98e08.r1 Soares fetal lung NbHL19W H...    126  0.9991
GB:C20625   C20625 HUMGS0004625, Human Gene Signature, 3'...    125  0.9997
GB:AA226777 AA226777 zr18d11.r1 Stratagene NT2 neuronal p...    125  0.9997
GB:AA128934 AA128934 zo08e09.r1 Stratagene neuroepitheliu...    125  0.9998
GB:VFLEB1   X14238 Vicia faba VfLEB1 pseudogene.               125  0.99993


>GB:AA282065 AA282065 zt02e02.s1 Soares NbHTGBC Homo sapiens cDNA clone 711962
3' similar to TR:G1167982 G1167982 ABC TRANSPORTER-7 ;.
Length = 450.

Minus Strand HSPs:

Score = 371 (102.5 bits), Expect = 1.2e-25, Sum P(2) = 1.2e-25
Identities = 77/81 (95%), Positives = 77/81 (95%), Strand = Minus / Plus

Query:   400 TTTNTTATGTGTGGCCCAAAGACAGGCCAGATATACGAGCTAGGGTTGCCATTTCCCTGG 341
             ||| ||||||||||||||||||||||||||| |||||||||| |||||||||| ||||
Sbjct:   320 TTTCTTATGTGTGGCCCAAAGACAGGCCAGATCTACGAGCTAGAGTTGCCATTTCGCTGG 379

Query:   340 GATTTTTGGGTGGTGCAAAGG 320
             |||||||||||||||||||||
Sbjct:   380 GATTTTTGGGTGGTGCAAAGG 400

Score = 125 (34.5 bits), Expect = 1.2e-25, Sum P(2) = 1.2e-25
Identities = 27/30 (90%), Positives = 27/30 (90%), Strand = Minus / Plus

Query:   443 GAACAGTTAAAAGATGTTGATANTCGGAAA 414
             |||  ||||||||||||||| |||||||
Sbjct:   274 GAAGGGTTAAAAGATGTTGATACTCGGAAA 303


>GB:MMU43892 U43892 Mus musculus ABC transporter-7 mRNA, partial cds.
Length = 2684

Minus Strand HSPs:

Score = 335 (92.6 bits), Expect = 2.2e-17, P = 2.2e-17
Identities = 73/81 (90%), Positives = 73/81 (90%), Strand = Minus / Plus

Query:   400 TTTNTTATGTGTGGCCCAAAGACAGGCCAGATATACGAGCTAGGGTTGCCATTTCCCTGG 341
             ||| ||||||||||||  |||| |||||||| ||||||| || ||||||||| |||||||
Sbjct:   200 TTTCTTATGTGTGGCCTGAAGATAGGCCAGATCTACGAGCCAGAGTTGCCATCTCCCTGG 259

Query:   340 GATTTTTGGGTGGTGCAAAGG 320
             |||||||||||||||||||||
Sbjct:   260 GATTTTTGGGTGGTGCAAAGG 280


>GB:MMU43892 U43892 Mus musculus ABC transporter-7 mRNA, partial cds.
Length = 2684

Minus Strand HSPs:

Score = 335 (92.6 bits), Expect = 2.2e-17, P = 2.2e-17
Identities = 73/81 (90%), Positives = 73/81 (90%), Strand = Minus / Plus

Query:   400 TTTNTTATGTGTGGCCCAAAGACAGGCCAGATATACGAGCTAGGGTTGCCATTTCCCTGG 341
             ||| ||||||||||||  |||| |||||||| ||||||| || ||||||||| |||||||
Sbjct:   200 TTTCTTATGTGTGGCCTGAAGATAGGCCAGATCTACGAGCCAGAGTTGCCATCTCCCTGG 259

Query:   340 GATTTTTGGGTGGTGCAAAGG 320
             |||||||||||||||||||||
Sbjct:   260 GATTTTTGGGTGGTGCAAAGG 280
```

> 5. Query= **x13**
(522 letters)

Sequences producing significant alignments:	Score (bits)	E Value
EM:HSU69729 U69729 Human Xp22 cosmid U110E1, complete sequence.	624	e-177
EM:AC008417 AC008417 Homo sapiens chromosome 5 clone CIT-HSPC_2...	44	0.019
EM:AC008417 AC008417 Homo sapiens chromosome 5 clone CIT-HSPC_2...	44	0.019
EM:HS433B8 Z93018 Human DNA sequence from PAC 433B8 on chromoso...	42	0.076
EM:AP000243 AP000243 Homo sapiens genomic DNA, chromosome 21q22...	42	0.076
EM:AP000204 AP000204 Homo sapiens genomic DNA, chromosome 21q22...	42	0.076
EM:AP000126 AP000126 Homo sapiens genomic DNA of 21q22.1, GART ...	42	0.076
EM:AC006256 AC006256 Homo sapiens chromosome 18, clone hRPK.12_...	42	0.076
EM:HS69B13 AL035698 Human DNA sequence from clone 69B13 on chro...	40	0.30
EM:AC006950 AC006950 Homo sapiens chromosome 19, cosmid R30669,...	40	0.30

>EM:HSU69729 U69729 Human Xp22 cosmid U110E1, complete sequence.
Length = 44185

Score = 624 bits (315), Expect = e-177
Identities = 413/442 (93%), Gaps = 4/442 (0%)
Strand = Plus / Plus

```
Query: 37   catctgtacaatgcagcaattgtttcatctttcccaaactaatagagatgctagaaacat 96
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct: 936  catctgtacaatgcagcaattgtttcatctttctcaaactaatagagatgctagaaacat 995


Query: 97   ctggtagttgatagttaattcatttttcattgcacctttcataggaagcagggaacgc 156
            |||||||||||||||||  ||||||||||||||||||||||||||||||||||| |
Sbjct: 996  ctggtagttgatagttaatttttttttcattgcacctttcataggaagcagggaacac 1055


Query: 157  aattcaatatatccccctgagagagtcttgcatttatgacaggtgtgtatttcttaacaag 216
            |||||||||| |||||||||| ||||||| ||||||||||||||||||||||||  ||
Sbjct: 1056 aattcaatatctcccctgagagggtcttgcgtttatgacaggtgtgtatttcttaatgag 1115


Query: 217  tcgttgtttgacaccattgcttctcctaagtccagctggaagtacattcatctggcaaaa 276
            |||||||||||||| ||||||||| ||||||||||||||||||||||||||||||||
Sbjct: 1116 tcgttgtttgacaccgttgcttctcttaagtccagctggaagtacattcatctggcaaaa 1175


Query: 277  agagagaggggaggaaaaatgaaatttgaaaggaagttggagtgtcaggtggttgttgca 336
            | |||||||||||||||| |||||||||||| ||||||||||||||||||||||||||
Sbjct: 1176 a-agagaggggaggaaaagtgaaatttgaaagaaagttggagtgtcaggtggttgttgca 1234


Query: 337  taaagcatccagttcactccagtaatatttattctgcttattttgtgcaatttcctgcaa 396
            |||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1235 taaaacatccaattcactccagtaatatttattctgcttattttgtgcaatttcctgcaa 1294


Query: 397  ctgccttttacaagtctgtcacaatgaagactttctctgttgtaagcgttatttaggtgc 456
            ||| ||||||||| |||||||| |||| |||||| |||||| |  | |||||||| |||
Sbjct: 1295 ctggcttttacaactctgtcactatgatgacttt-tctgtttt--gagttatttagctgc 1351


Query: 457  atactactaatttatttgagac 478
              ||||| |||||||| ||||||
Sbjct: 1352 gtactattaatttatctgagac 1373
```

6. Query= **x56**
(657 letters)

Sequences producing significant alignments:	Score (bits)	E Value
EM:CF33628 U33628 Canis familiaris sphingolipid Ca2+ release me...	81	8e-15
EM:U89607 U89607 Canis familiaris chymase gene, complete cds.	77	1e-13
EM:CFZUBECA2 AJ003059 Canis familiaris microsatellite DNA, ZuBe...	77	1e-13
EM:CDCOLIP M63427 Dog pancreatic colipase gene, complete cds.	75	5e-13
EM:AF130680 AF130680 Felis catus clone Fca641 microsatellite se...	75	5e-13
EM:CFA239533 AJ239533 Canis familiaris tRNA-derived SINE elemen...	73	2e-12
EM:CFA388552 AJ388552 Canis familiaris mRNA for partial hypothe...	71	8e-12
EM:AF029683 AF029683 Canis familiaris tyrosinase-related protei...	71	8e-12
EM:CFA239547 AJ239547 Canis familiaris tRNA-derived SINE elemen...	69	3e-11
EM:CFA239549 AJ239549 Canis familiaris tRNA-derived SINE elemen...	67	1e-10

>EM:CF33628 U33628 Canis familiaris sphingolipid Ca2+ release mediating protein of endoplasmic reticulum mRNA, complete cds.
Length = 1869

Score = 80.5 bits (40), Expect = 8e-15
Identities = 56/63 (88%)
Strand = Plus / Minus

```
Query: 475  aggcngaggganancaggctccntgcagggaacccgacntgggactcgatcccgggtct 534
            |||| |||||| || |||||||| |||||||| |||||| |||||||| ||||||||||
Sbjct: 1236 aggcagagggagaagcaggctccatgcagggagcccgacatgggactccatcccgggtct 1177


Query: 535  cca 537
            |||
Sbjct: 1176 cca 1174
```

Score = 34.5 bits (17), Expect = 0.55
Identities = 24/25 (96%), Gaps = 1/25 (4%)
Strand = Plus / Minus

```
Query: 429  aagattttat-tatttattcatgag 452
            |||||||||| ||||||||||||||
Sbjct: 1269 aagattttatgtatttattcatgag 1245
```

>EM:U89607 U89607 Canis familiaris chymase gene, complete cds.
Length = 8747

Score = 76.5 bits (38), Expect = 1e-13
Identities = 69/80 (86%), Gaps = 1/80 (1%)
Strand = Plus / Minus

```
Query: 475  aggcngaggganancaggctccntgcagggaacccgacntgggactcgatcccgggtct 534
            |||| |||||| || |||||||| |||||||| || ||| ||||||||||||||| |||||
Sbjct: 4121 aggcagagggagaagcaggctccatgcagggagcctgacgtgggactcgatcccaggtct 4062


Query: 535  cca-natcatgccctgggcc 553
            |||  |||| ||||||||||
Sbjct: 4061 ccaggatcacgccctgggcc 4042
```

Score = 54.5 bits (27), Expect = 5e-07
Identities = 50/58 (86%), Gaps = 1/58 (1%)
Strand = Plus / Minus

```
Query: 475 aggcngaggganancaggctccntgcagggaacccgacntggg-actcgatcccggg 531
           |||| |||||| || |||||||| |||||||| |||||  |||| ||||||||||||
Sbjct: 393 aggcagagggagaagcaggctccatgcagggagcccgatgtggggactcgatcccggg 336
```

Score = 52.5 bits (26), Expect = 2e-06
Identities = 38/42 (90%)
Strand = Plus / Plus

```
Query: 257  caggaccttgagatcatgacctgagcagaaatcaagagtcag 298
```

```
            ||||||||| ||||||||||  |||||||| ||||||||||
Sbjct: 3037 caggacctttagatcatgacccaagcagaaaccaagagtcag 3078


 Score = 42.5 bits (21), Expect = 0.002
 Identities = 42/49 (85%)
 Strand = Plus / Plus


Query: 283  agaaatcaagagtcagaggctccaccgactgagccacccaagcacccct 331
            ||||| ||||||||||   ||| ||| ||||||||||||| ||||||||
Sbjct: 3282 agaaaccaagagtcagccccctcaaccaactgagccacccaggcacccct 3330


 Score = 34.5 bits (17), Expect = 0.55
 Identities = 20/21 (95%)
 Strand = Plus / Minus


Query: 195  gcagagggagagggagaagca 215
            ||||||| |||||||||||||
Sbjct: 3729 gcagaggaagagggagaagca 3709


 Score = 34.5 bits (17), Expect = 0.55
 Identities = 24/25 (96%), Gaps = 1/25 (4%)
 Strand = Plus / Minus


Query: 429  aagattttatt-atttattcatgag 452
            |||||||||| ||||||||||||
Sbjct: 4178 aagattttatttatttattcatgag 4154


 Score = 32.5 bits (16), Expect = 2.2
 Identities = 23/24 (95%), Gaps = 1/24 (4%)
 Strand = Plus / Minus


Query: 429  aagattttatt-atttattcatga 451
            |||||||||| ||||||||||||
Sbjct: 3243 aagattttatttatttattcatga 3220


 Score = 30.5 bits (15), Expect = 8.8
 Identities = 15/15 (100%)
 Strand = Plus / Minus


Query: 438  ttatttattcatgag 452
            |||||||||||||||
Sbjct: 5075 ttatttattcatgag 5061


>EM:CFZUBECA2 AJ003059 Canis familiaris microsatellite DNA, ZuBeCa2
          Length = 589

 Score = 76.5 bits (38), Expect = 1e-13
 Identities = 69/80 (86%), Gaps = 1/80 (1%)
 Strand = Plus / Plus


Query: 475 aggcngaggganaancaggctccntgcagggaacccgacntgggactcgatcccgggtct 534
           |||| |||||| || |||||||| ||| |||| |||||| ||||||||||||||||| ||
Sbjct: 435 aggcagagggagaagcaggctccatgctgggagcccgacgtgggactcgatcccgggact 494


Query: 535 cca-natcatgccctgggcc 553
           ||| ||| ||||||||||||
Sbjct: 495 ccaggatcgtgccctgggcc 514


>EM:CDCOLIP M63427 Dog pancreatic colipase gene, complete cds.
          Length = 3164

 Score = 74.5 bits (37), Expect = 5e-13
 Identities = 63/72 (87%), Gaps = 1/72 (1%)
 Strand = Plus / Plus
```

```
Query: 482  ggganaancaggctccntgcagggaacccgacntgggactcgatcccgggtctcca-nat 540
            |||| || |||||||| |||||||| || ||| |||||||||||||| |||||||||  ||
Sbjct: 1229 gggagaagcaggctccatgcagggagcctgacgtgggactcgatcctgggtctccatgat 1288


Query: 541  catgccctgggc 552
            ||||||||||||
Sbjct: 1289 catgccctgggc 1300


 Score = 46.5 bits (23), Expect = 1e-04
 Identities = 54/65 (83%), Gaps = 2/65 (3%)
 Strand = Plus / Minus


Query: 475 aggcngagggganaancaggctccntgca--gggaacccgacntgggactcgatcccgggt 532
            |||| |||||| || |||||||| ||||  |||| || ||  ||||| ||||||||||||
Sbjct: 628 aggcagagggagaagcaggctccatgcaccgggagcctgatgtgggattcgatcccgggt 569


Query: 533 ctcca 537
            |||||
Sbjct: 568 ctcca 564


 Score = 32.5 bits (16), Expect = 2.2
 Identities = 23/24 (95%), Gaps = 1/24 (4%)
 Strand = Plus / Minus


Query: 429 aagattttatt-atttattcatga 451
            |||||||||||| ||||||||||||
Sbjct: 685 aagattttatttatttattcatga 662


>EM:AF130680 AF130680 Felis catus clone Fca641 microsatellite
           sequence.
           Length = 259

 Score = 74.5 bits (37), Expect = 5e-13
 Identities = 58/65 (89%)
 Strand = Plus / Minus


Query: 268 gatcatgacctgagcagaaatcaagagtcagaggctccaccgactgagccacccaagcac 327
            |||||||||||||||| |||||||||||||||  ||   |||||||||||||||| || |
Sbjct: 259 gatcatgacctgagctgaaatcaagagtcagatgccgaaccgactgagccacccaggcgc 200


Query: 328 cccta 332
            |||||
Sbjct: 199 cccta 195


>EM:CFA239533 AJ239533 Canis familiaris tRNA-derived SINE element,
           clone D254TB5
           Length = 676

 Score = 72.5 bits (36), Expect = 2e-12
 Identities = 55/63 (87%)
 Strand = Plus / Minus


Query: 475 aggcngagggganaancaggctccntgcagggaacccgacntgggactcgatcccgggtct 534
            |||| |||||| || |||||||| |||||||| |||||| ||||||| ||| ||||||||
Sbjct: 471 aggcagagggagaagcaggctccatgcagggagcccgacgtgggacttgattccgggtct 412


Query: 535 cca 537
            |||
Sbjct: 411 cca 409


 Score = 34.5 bits (17), Expect = 0.55
 Identities = 17/17 (100%)
 Strand = Plus / Minus


Query: 306 accgactgagccaccca 322
```

```
            ||||||||||||||||||
Sbjct: 655 accgactgagccaccca 639


>EM:CFA388552 AJ388552 Canis familiaris mRNA for partial
           hypothetical protein, clone V2.89
           Length = 893

 Score = 70.5 bits (35), Expect = 8e-12
 Identities = 54/62 (87%)
 Strand = Plus / Minus


Query: 475 aggcngaggganaancaggctccntgcagggaacccgacntgggactcgatcccgggtct 534
           |||| ||||||| || ||||||||  ||||||| |||||| |||| ||||||||||||||
Sbjct: 804 aggcagagggagaagcaggctccaggcagggagcccgacgtggggctcgatcccgggtct 745


Query: 535 cc 536
           ||
Sbjct: 744 cc 743


>EM:AF029683 AF029683 Canis familiaris tyrosinase-related protein 2
           gene, exons 2 and 3 and partial cds.
           Length = 2850

 Score = 70.5 bits (35), Expect = 8e-12
 Identities = 57/65 (87%), Gaps = 2/65 (3%)
 Strand = Plus / Minus


Query: 475 aggcngaggganaancaggctccntgca--gggaacccgacntgggactcgatcccgggt 532
           |||| |||||| || |||||||| ||||  |||| |||||| ||||||||||||||||||
Sbjct: 805 aggcagagggagaagcaggctccatgcaccgggagcccgacgtgggactcgatcccgggt 746


Query: 533 ctcca 537
           |||||
Sbjct: 745 ctcca 741


 Score = 34.5 bits (17), Expect = 0.55
 Identities = 24/25 (96%), Gaps = 1/25 (4%)
 Strand = Plus / Plus


Query: 429  aagattttatt-atttattcatgag 452
             |||||||||||| ||||||||||||
Sbjct: 1179 aagattttatttatttattcatgag 1203


 Score = 30.5 bits (15), Expect = 8.8
 Identities = 15/15 (100%)
 Strand = Plus / Plus


Query: 268  gatcatgacctgagc 282
             |||||||||||||||
Sbjct: 1292 gatcatgacctgagc 1306


>EM:CFA239547 AJ239547 Canis familiaris tRNA-derived SINE element,
           clone D254ZG3
           Length = 674

 Score = 68.5 bits (34), Expect = 3e-11
 Identities = 50/57 (87%)
 Strand = Plus / Minus


Query: 475 aggcngaggganaancaggctccntgcagggaacccgacntgggactcgatcccggg 531
           |||| ||||||| || |||||||| |||||||| |||||  ||||||||||||||||
Sbjct: 575 aggcagagggagaagcaggctccatgcagggagcccgatgtgggactcgatcccggg 519


 Score = 34.5 bits (17), Expect = 0.55
 Identities = 24/25 (96%), Gaps = 1/25 (4%)
 Strand = Plus / Minus
```

```
Query: 429 aagattttatt-atttattcatgag 452
           ||||||||||| ||||||||||||
Sbjct: 644 aagattttatttatttattcatgag 620
```

```
>EM:CFA239549 AJ239549 Canis familiaris tRNA-derived SINE element,
           clone D261YA9
           Length = 679

 Score = 66.5 bits (33), Expect = 1e-10
 Identities = 50/57 (87%)
 Strand = Plus / Minus
```

```
Query: 480 gaggganaancaggctccntgcagggaacccgacntgggactcgatcccgggtctcc 536
           |||||| || |||||||| |||| |||||||||| |||||||||| || |||||||
Sbjct: 142 gagggagaagcaggctccatgcatggaacccgacgtgggactcgataccaggtctcc 86
```

```
 Score = 42.5 bits (21), Expect = 0.002
 Identities = 21/21 (100%)
 Strand = Plus / Minus
```

```
Query: 195 gcagagggagagggagaagca 215
           |||||||||||||||||||||
Sbjct: 151 gcagagggagagggagaagca 131
```

```
 Score = 34.5 bits (17), Expect = 0.55
 Identities = 24/25 (96%), Gaps = 1/25 (4%)
 Strand = Plus / Minus
```

```
Query: 429 aagattttatt-atttattcatgag 452
           ||||||||||| ||||||||||||
Sbjct: 206 aagattttatttatttattcatgag 182
```

Appendix 4

Table iii Genotyping results with chromosome X-derived markers

The table shows the data in the form required by the linkage analysis software.

The first column gives the names of the members of the Cornell families from family 4 to family 7.

The linkage data is shown in the next 9 columns:

Column 1 Pedigree number, designated 1 - 4 for family 4 - 7 respectively

Column 2 Individual identity number

Column 3 Identity number of father

Column 4 Identity number of mother

Column 5 First offspring Identity number

Column 6 Next paternal sibling identity number

Column 7 Next maternal sibling identity number

Column 8 Sex (where 1 = male and 2 = female)

Column 9 Proband and loop break status

The typing results for each marker are shown with the alleles given as integers (rather than actual PCR product sizes).

The final column has the individuals identity within each family as used by the linkage programme.

Table iii

Dog	Linkage Data									X 5 8	F 9	x 6 6	X 9	F 8 c	X 2 7	X 6 8	H 6 2	X 5 0	G K 1	X 2 4	X 6 4	X 2 8	X 1 2	X 5 1	x 2 0	X 2 1	
CF4-2	1	1	0	0	3	0	0	1	1	22	11	00	00	22	55	22	00	11	00	00	11	37	22	11	13	00	Per: 1
CF4-1	1	2	0	0	3	0	0	2	0	12	11	00	00	23	15	12	00	22	00	00	23	36	22	13	22	00	Per: 2
CF4-5	1	3	1	2	7	0	0	2	0	12	11	12	22	23	15	12	00	12	00	00	12	67	22	00	12	00	Per: 3
CF4-6	1	4	5	6	7	0	0	1	0	33	22	11	00	33	15	22	00	22	00	00	11	37	12	33	12	00	Per: 4
CF4-4	1	5	0	0	4	0	0	1	0	33	22	00	00	33	13	22	00	22	00	00	11	23	11	33	22	00	Per: 5
CF4-3	1	6	0	0	4	0	0	2	0	23	12	00	00	23	55	22	00	12	00	00	11	37	12	13	12	00	Per: 6
CF4-7	1	7	4	3	0	8	8	1	0	11	11	11	00	33	15	11	00	22	00	00	11	37	22	13	22	00	Per: 7
CF4-8	1	8	4	3	0	9	9	1	0	22	11	11	00	33	15	22	00	11	00	00	11	67	22	13	22	00	Per: 8
CF4-9	1	9	4	3	0	10	10	1	0	22	11	11	00	33	55	22	00	11	00	00	11	37	12	13	11	00	Per: 9
CF4-10	1	10	4	3	0	11	11	1	0	22	11	11	00	22	15	22	00	11	00	00	11	37	22	13	22	00	Per: 10
CF4-11	1	11	4	3	0	12	12	2	0	13	12	12	00	23	15	12	00	22	00	00	11	67	12	13	11	00	Per: 11
CF4-12	1	12	4	3	0	13	13	2	0	23	12	12	00	23	15	22	00	22	00	00	11	37	00	13	22	00	Per: 12
CF4-13	1	13	4	3	0	14	14	1	0	11	11	11	00	33	55	11	00	11	00	00	11	67	12	13	22	00	Per: 13
CF4-14	1	14	4	3	0	15	15	1	0	22	11	00	22	33	15	22	00	11	00	00	11	37	12	13	22	00	Per: 14
CF4-15	1	15	4	3	0	16	16	1	0	22	11	11	00	33	15	22	00	11	00	00	11	36	12	13	22	00	Per: 15
CF4-16	1	16	4	3	0	17	17	1	0	11	11	11	00	33	55	11	00	22	00	00	22	36	22	13	11	00	Per: 16
CF4-17	1	17	4	3	0	18	18	2	0	13	12	12	00	00	15	12	00	22	00	00	11	77	22	13	22	00	Per: 17
CF4-18	1	18	4	3	0	19	19	2	0	13	12	12	00	00	15	12	00	22	00	00	11	36	12	13	11	00	Per: 18
CF4-19	1	19	4	3	0	20	20	2	0	13	12	12	00	00	15	22	00	22	00	00	11	77	12	13	22	00	Per: 19
CF4-20	1	20	4	3	0	21	21	2	0	13	12	12	00	00	55	12	00	22	00	00	11	67	22	13	22	00	Per: 20
CF4-21	1	21	4	3	0	22	22	2	0	23	12	12	00	23	15	12	00	12	00	00	11	37	12	13	22	00	Per: 21
CF4-22	1	22	4	3	0	0	0	2	0	23	12	12	00	23	15	22	00	12	00	00	12	37	22	13	22	00	Per: 22
CF5-1	2	1	0	0	3	0	0	1	1	00	00	00	11	11	25	22	00	00	00	00	00	36	00	12	11	00	Per: 1
CF5-2	2	2	0	0	3	0	0	2	0	00	00	00	00	33	34	12	00	00	00	00	00	23	00	33	22	00	Per: 2
CF5-7	2	3	1	2	7	0	0	1	0	00	00	00	11	33	24	22	00	00	00	22	00	33	00	23	12	00	Per: 3
CF5-8	2	4	5	6	7	0	0	2	0	00	00	12	12	34	14	12	00	22	00	12	00	33	00	13	22	00	Per: 4
CF5-4	2	5	0	0	4	0	0	1	0	00	22	00	11	33	44	22	00	00	00	22	00	37	00	13	22	00	Per: 5
CF5-3	2	6	0	0	4	0	0	2	0	00	00	00	12	34	11	12	00	00	00	12	00	33	00	13	22	00	Per: 6
CF5-10	2	7	3	4	0	8	8	1	0	00	00	12	11	33	14	22	00	00	00	00	00	33	00	23	22	00	Per: 7
CF5-11	2	8	3	4	0	9	9	1	0	00	00	12	11	33	14	22	00	00	00	00	00	33	00	13	22	00	Per: 8
CF5-12	2	9	3	4	0	10	10	2	0	00	00	00	12	33	44	12	00	00	00	00	00	33	00	33	12	00	Per: 9
CF5-13	2	10	3	4	0	11	11	2	0	00	00	12	12	33	12	12	00	00	00	00	00	33	00	23	22	00	Per: 10
CF5-14	2	11	3	4	0	12	12	2	0	00	00	12	12	33	44	12	00	00	00	00	00	33	00	12	22	00	Per: 11
CF5-15	2	12	3	4	0	0	16	2	0	00	00	12	12	33	14	12	00	00	00	00	00	33	00	13	22	00	Per: 12
CF5-9	2	13	14	15	16	0	0	1	0	00	00	00	22	11	15	11	00	11	00	11	00	56	00	00	22	00	Per: 13
CF5-6	2	14	0	0	13	0	0	1	0	00	22	00	22	33	15	22	00	00	00	22	00	36	00	33	22	00	Per: 14
CF6-1	2	15	0	0	13	0	0	2	0	00	11	33	12	12	11	11	00	00	00	12	00	15	00	11	22	00	Per: 15
CF5-16	2	16	13	4	0	17	17	1	0	00	00	12	00	33	11	22	00	00	00	22	00	35	00	13	22	00	Per: 16
CF5-17	2	17	13	4	0	18	18	1	0	00	00	12	11	33	11	22	00	00	00	22	00	36	00	13	22	00	Per: 17
CF5-18	2	18	13	4	0	19	19	2	0	00	00	13	00	13	15	12	00	00	00	00	00	36	00	00	22	00	Per: 18
CF5-19	2	19	13	4	0	20	20	1	0	00	00	11	11	33	11	22	00	00	00	22	00	36	00	13	22	00	Per: 19
CF5-20	2	20	13	4	0	21	21	2	0	00	00	11	12	13	14	12	00	00	00	12	00	36	00	11	22	00	Per: 20
CF5-21	2	21	13	4	0	22	22	2	0	00	00	11	12	13	00	12	00	00	00	11	00	35	00	13	22	00	Per: 21
CF5-22	2	22	13	4	0	23	23	2	0	00	00	11	12	13	14	12	00	00	00	12	00	35	00	00	22	00	Per: 22
CF5-23	2	23	13	4	0	24	24	2	0	00	00	11	12	13	15	12	00	00	00	12	00	35	00	13	22	00	Per: 23
CF5-24	2	24	13	4	0	0	0	2	0	00	00	11	12	13	11	12	00	00	00	12	00	00	00	00	22	00	Per: 24
CF4-6	3	1	0	0	3	0	0	1	1	33	22	00	11	00	00	00	22	22	00	00	00	37	12	00	12	23	Per: 1
CF4-5	3	2	0	0	3	0	0	2	0	12	11	00	22	00	00	00	23	12	00	00	00	67	22	00	12	33	Per: 2
CF4-14	3	3	1	2	7	0	0	1	0	22	11	00	22	00	00	00	33	11	00	00	00	37	12	00	22	23	Per: 3
CF6-2	3	4	5	6	7	0	0	2	0	23	12	00	12	00	00	00	12	12	00	00	00	35	12	00	00	13	Per: 4
CF5-5	3	5	0	0	4	0	0	1	0	33	22	00	22	00	00	00	22	22	00	00	00	36	11	00	22	23	Per: 5
CF6-1	3	6	0	0	4	0	0	2	0	00	11	00	12	00	00	00	13	12	00	00	00	15	22	00	22	13	Per: 6
CF6-3	3	7	3	4	0	8	8	1	0	22	11	00	11	00	00	00	11	11	00	00	00	35	22	00	00	13	Per: 7
CF6-4	3	8	3	4	0	9	9	1	0	22	11	00	11	00	00	00	11	11	00	00	00	35	12	00	00	33	Per: 8
CF6-7	3	9	3	4	0	10	10	1	0	22	11	00	11	00	00	00	11	11	00	00	00	35	12	00	00	13	Per: 9
CF6-5	3	10	3	4	0	11	11	1	0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	Per: 10
CF6-6	3	11	3	4	0	12	12	2	0	23	12	00	22	00	00	00	00	00	00	00	00	33	12	00	00	33	Per: 11
CF6-8	3	12	3	4	0	13	13	2	0	23	12	00	22	00	00	00	00	00	00	00	00	35	12	00	00	13	Per: 12
CF6-9	3	13	3	4	0	14	14	1	0	33	22	00	22	00	00	00	22	22	00	00	00	37	00	00	00	33	Per: 13
CF6-10	3	14	3	4	0	15	15	1	0	22	11	00	22	00	00	00	11	11	00	00	00	35	12	00	00	13	Per: 14
CF6-11	3	15	3	4	0	16	16	1	0	22	11	00	11	00	00	00	11	11	00	00	00	37	11	00	00	13	Per: 15
CF6-12	3	16	3	4	0	17	17	1	0	33	22	00	11	00	00	00	11	11	00	00	00	33	22	00	00	00	Per: 16
CF6-13	3	17	3	4	0	18	18	1	0	22	11	00	11	00	00	00	11	11	00	00	00	37	11	00	00	00	Per: 17
CF6-14	3	18	3	4	0	19	19	2	0	22	11	00	00	00	00	00	00	00	00	00	00	33	12	00	00	13	Per: 18
CF6-15	3	19	3	4	0	0	23	2	0	23	11	00	12	00	00	00	13	00	00	00	00	33	12	00	00	00	Per: 19
CF4-6	3	20	21	22	23	0	0	1	0	33	22	00	11	00	00	00	22	22	00	00	00	37	12	00	12	23	Per: 20
CF4-4	3	21	0	0	20	0	0	1	0	33	22	00	11	00	00	00	22	22	00	00	00	23	11	00	22	23	Per: 21
CF4-3	3	22	0	0	20	0	0	2	0	23	12	00	12	00	00	00	23	12	00	00	00	37	12	00	12	23	Per: 22
CF6-16	3	23	20	4	0	24	24	1	0	33	22	00	22	00	00	00	22	22	00	00	00	33	12	00	00	00	Per: 23
CF6-17	3	24	20	4	0	25	25	1	0	22	11	00	22	00	00	00	22	22	00	00	00	33	12	00	00	13	Per: 24

Table iii

Dog	Linkage Data									X58	F9	x66	X9	F8c	X27	X68	H62	X50	G K1	X24	X64	X28	X12	X51	x20	X21		
CF6-18	3	25	20	4	0	26	26	1	0	22	11	00	22	00	00	00	22	22	00	00	00	33	12	00	00	33	Per:	25
CF6-19	3	26	20	4	0	27	27	1	0	33	22	00	11	00	00	00	22	11	00	00	00	57	22	00	00	33	Per:	26
CF6-20	3	27	20	4	0	28	28	2	0	23	12	00	11	00	00	00	12	00	00	00	00	35	12	00	00	13	Per:	27
CF6-21	3	28	20	4	0	29	29	2	0	33	22	00	12	00	00	00	22	00	00	00	00	57	22	00	00	33	Per:	28
CF6-23	3	29	20	4	0	30	30	2	0	23	12	00	11	00	00	00	22	00	00	00	00	57	12	00	00	33	Per:	29
CF6-22	3	30	20	4	0	0	0	2	0	23	12	00	12	00	00	00	22	00	00	00	00	35	22	00	00	33	Per:	30
CF4-14	4	1	0	0	3	0	0	1	1	22	11	00	22	33	00	00	33	11	22	00	00	37	12	00	22	00	Per:	1
CF5-8	4	2	0	0	3	0	0	2	0	13	22	12	12	34	00	00	22	22	11	00	00	33	22	00	22	00	Per:	2
CF7-3	4	3	1	2	7	0	0	2	0	23	12	12	12	33	00	00	23	12	12	00	00	37	12	00	12	00	Per:	3
CF7-4	4	4	5	6	7	0	0	1	0	33	22	11	11	44	00	00	22	22	11	00	00	35	12	00	12	00	Per:	4
CF1-6	4	5	0	0	4	0	0	1	0	00	11	00	22	22	00	00	33	22	11	00	00	15	12	00	00	00	Per:	5
CF7-1	4	6	0	0	4	0	0	2	0	00	12	00	12	34	00	00	23	22	11	00	00	33	11	00	00	00	Per:	6
CF7-7	4	7	4	3	0	8	8	1	0	22	11	11	11	00	00	00	33	11	22	00	00	35	12	00	00	00	Per:	7
CF7-8	4	8	4	3	0	9	9	1	0	22	22	11	11	33	00	00	22	11	22	00	00	35	22	00	00	00	Per:	8
CF7-9	4	9	4	3	0	10	10	1	0	22	11	11	22	00	00	00	33	11	22	00	00	57	22	00	00	00	Per:	9
CF7-10	4	10	4	3	0	11	11	2	0	33	22	12	11	00	00	00	22	22	11	00	00	35	12	00	00	00	Per:	10
CF7-11	4	11	4	3	0	12	12	2	0	23	12	11	12	00	00	00	23	12	12	00	00	00	22	00	00	00	Per:	11
CF7-12	4	12	4	3	0	16	0	2	0	23	12	11	12	00	00	00	23	12	12	00	00	00	11	00	00	00	Per:	12
CF5-8	4	13	14	15	16	0	0	2	0	13	22	00	12	34	00	00	22	22	11	00	00	33	22	00	22	00	Per:	13
CF5-4	4	14	0	0	13	0	0	1	0	33	22	00	11	33	00	00	00	22	00	00	00	37	00	00	22	00	Per:	14
CF5-3	4	15	0	0	13	0	0	2	0	13	12	00	12	34	00	00	00	22	00	00	00	33	00	00	22	00	Per:	15
CF7-13	4	16	4	13	0	17	17	1	0	00	00	00	11	44	00	00	00	00	00	00	00	35	00	00	00	00	Per:	16
CF7-14	4	17	4	13	0	18	18	1	0	00	00	00	11	44	00	00	00	00	00	00	00	35	00	00	00	00	Per:	17
CF7-15	4	18	4	13	0	19	19	1	0	00	00	00	11	44	00	00	00	00	00	00	00	35	00	00	00	00	Per:	18
CF7-16	4	19	4	13	0	20	20	1	0	00	00	00	11	44	00	00	00	00	00	00	00	33	00	00	00	00	Per:	19
CF7-17	4	20	4	13	0	21	21	1	0	00	00	00	11	44	00	00	00	00	00	00	00	35	00	00	00	00	Per:	20
CF7-18	4	21	4	13	0	22	22	2	0	00	00	00	11	34	00	00	00	00	00	00	00	33	00	00	00	00	Per:	21
CF7-19	4	22	4	13	0	26	0	2	0	00	00	00	11	34	00	00	00	00	00	00	00	35	00	00	00	00	Per:	22
CF7-5	4	23	24	25	26	0	0	2	0	23	12	00	12	33	00	00	23	12	12	00	00	33	12	00	12	00	Per:	23
CF4-14	4	24	0	0	23	0	0	1	0	22	11	00	22	33	00	00	33	11	22	00	00	37	12	00	22	00	Per:	24
CF5-15	4	25	0	0	23	0	0	2	0	13	12	00	12	33	00	00	23	22	12	00	00	33	12	00	22	00	Per:	25
CF7-25	4	26	4	23	0	27	27	1	0	33	22	00	11	00	00	00	22	22	11	00	00	33	12	00	00	00	Per:	26
CF7-26	4	27	4	23	0	28	28	1	0	22	11	00	22	00	00	00	33	11	22	00	00	35	12	00	00	00	Per:	27
CF7-27	4	28	4	23	0	29	29	1	0	22	11	00	22	00	00	00	33	11	22	00	00	33	12	00	00	00	Per:	28
CF7-28	4	29	4	23	0	30	30	1	0	22	11	00	22	00	00	00	33	11	22	00	00	33	12	00	00	00	Per:	29
CF7-29	4	30	4	23	0	31	31	1	0	22	11	00	00	00	00	00	33	11	11	00	00	35	12	00	00	00	Per:	30
CF7-30	4	31	4	23	0	32	32	1	0	22	11	00	22	00	00	00	33	11	22	00	00	35	12	00	00	00	Per:	31
CF7-31	4	32	4	23	0	33	33	2	0	23	22	00	12	00	00	00	23	12	12	00	00	00	11	00	00	00	Per:	32
CF7-20	4	33	4	23	0	34	34	1	0	33	22	00	11	00	00	00	22	22	11	00	00	35	22	00	00	00	Per:	33
CF7-21	4	34	4	23	0	35	35	1	0	22	11	00	22	00	00	00	33	11	22	00	00	33	11	00	00	00	Per:	34
CF7-22	4	35	4	23	0	36	36	2	0	23	12	00	12	00	00	00	23	12	12	00	00	35	22	00	00	00	Per:	35
CF7-23	4	36	4	23	0	37	37	2	0	23	12	00	12	00	00	00	23	12	12	00	00	35	12	00	00	00	Per:	36
CF7-24	4	37	4	23	0	0	0	2	0	33	12	00	00	00	00	00	23	22	11	00	00	35	00	00	00	00	Per:	37

Appendix 5

Table iv Typing results of WG-RH panel with chromosome X-derived markers

Table v Typing results of WG-RH panel with chromosome 1-derived markers

For each table, the cell line names and numbers are given on the left (as given by Research Genetics).

Dog = dog genomic DNA as a positive control, A23 = DNA from hamster cell line A23 used to produce the WG-RH panel, MQ = MQ water as negative control.

The top-most row shows the marker name.

"+" indicates a PCR product was produced, "-" indicates no product observed and "?" indicates an ambiguous result.

The penultimate row shows the retention frequency for each marker, the last row gives the average figure for each chromosome.

Table iv

CELL LINE	HYBRID	x91	x78r	x77	x74	x71	x89	x88	x87	x86	x84	x58	x57	x51	x50	x49	x47	x40	x35	x34	x33	x31	x30	x28	x27	x24	x21	x20	x13	x12	x9	x1	PGK	F8c	F9	PDHA2	1E7	CHM
A1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A3	2	-	-	-	-	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A6	3	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	-	+	-
B3	4	-	-	+	-	-	-	-	+	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-
B4	5	-	-	+	-	-	-	+	-	-	+	+	+	-	-	+	-	-	-	-	-	-	+	-	-	+	-	+	-	-	?	-	+	+	+	-	-	+
B5	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
C2	7	+	-	-	+	+	+	-	+	+	+	-	+	-	+	+	-	+	-	-	+	+	-	-	-	-	+	-	+	-	-	-	-	-	-	+	+	-
C3	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E2	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
E6	10	-	-	-	-	-	+	+	-	+	-	+	+	-	+	+	-	+	-	+	+	-	+	+	-	-	-	-	-	+	+	+	-	+	+	-	-	+
F1	11	-	-	+	+	+	+	+	-	+	+	+	+	-	+	+	-	+	+	+	+	-	+	+	+	+	-	+	-	-	+	+	+	+	+	-	-	+
G1	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-
H6	13	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-
I2	14	+	+	+	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	+	-	-	+	+	+
I4	15	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-
I5	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I6	17	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	-	-	-	-	-	+	+	-	+	+	+	-	-	-	-	-	-	-	+	-
J3	18	+	+	-	-	-	-	+	-	-	-	+	-	+	-	+	-	-	-	-	-	+	+	+	-	-	+	-	+	-	-	-	-	+	+	+	+	-
K1	19	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-
L3	20	-	-	-	-	-	+	+	-	+	+	+	+	-	+	+	-	+	+	-	+	-	+	-	-	-	-	-	-	-	+	+	-	-	+	-	-	+
L4	21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M1	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-
M3	23	+	+	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	-	+	-	+	+	-	-	+	-	+	-	-	+	-	-	-	-	+	-	-
N1	24	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
N3	25	+	+	-	-	-	+	-	+	+	+	+	-	-	+	+	+	+	-	-	+	+	-	+	-	+	+	-	+	-	-	-	-	-	-	+	+	-
N5	26	-	-	-	+	-	-	-	+	-	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-
N6	27	-	-	-	+	+	-	-	+	-	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-
P1	28	+	-	-	-	-	+	-	-	+	-	-	-	+	-	-	-	-	+	+	+	-	+	-	+	+	-	-	-	-	+	-	-	-	-	+	-	-
P2	29	-	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Q1	30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+
Q2	31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Q4	32	+	+	-	+	-	-	-	?	-	-	-	+	-	-	-	-	-	-	+	+	+	+	-	-	-	-	+	+	-	+	-	-	-	-	+	-	-
Q5	33	-	-	-	+	+	-	+	+	-	+	+	-	-	-	+	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	-	+	-	-	+
Q6	34	-	-	-	+	-	+	+	-	-	+	+	-	-	+	+	+	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	+	-	+	-	-	-
R1	35	-	-	-	+	+	-	-	-	-	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-
R2	36	-	-	+	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-
R3	37	-	-	-	+	+	-	-	+	-	+	-	+	+	-	+	+	-	-	-	-	-	+	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-
S1	38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-
S2	39	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-
S4	40	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-
T5	41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
T6	42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
U2	43	-	-	+	-	+	+	-	-	+	+	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	+	+
V1	44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
V3	45	+	+	-	-	-	-	-	+	-	+	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+
V4	46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V5	47	-	-	-	-	-	+	+	-	-	-	+	-	-	+	-	+	+	-	-	+	-	-	+	+	-	-	-	-	+	-	-	-	+	+	-	?	-
W3	48	+	-	+	-	-	+	+	-	+	+	-	+	-	-	+	-	+	+	-	+	-	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	+
W4	49	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-
W5	50	-	-	+	+	+	+	-	+	+	+	-	-	+	+	+	+	+	-	-	+	-	-	+	+	+	-	+	-	-	-	-	+	-	-	-	-	+

Table Iv

CELL LINE	HYBRID	x91	x78r	x77	x74	x71	x69	x68	x67	x66	x64	x58	x57	x51	x50	x49	x47	x40	x35	x34	x33	x31	x30	x28	x27	x24	x21	x20	x13	x12	x9	x1	PGK	F8c	F9	PDHA2	1E7	CHM
W6	51	-	-	+	+	+	+	-	+	+	+	-	-	+	+	+	-	+	-	-	+	-	-	+	+	-	-	-	-	-	-	+	-	-	-	-	-	+
X1	52	+	-	+	-	-	+	-	-	+	+	-	+	-	-	+	-	+	+	-	+	-	+	+	+	-	-	+	-	-	+	+	-	-	+	-	+	
X6	53	-	-	-	-	-	-	+	-	-	+	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	
Y2	54	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Y3	55	-	-	-	-	-	+	-	-	+	-	+	-	-	-	-	-	+	+	-	+	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-
Z1	56	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Z3	57	-	-	-	-	-	-	+	-	-	+	-	-	+	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Z5	58	-	-	-	-	-	+	+	+	+	+	+	+	+	-	+	-	+	+	-	+	-	-	-	+	+	-	+	-	-	-	-	+	-	+	-	-	-
2A1	59	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2A3	60	-	-	-	-	+	-	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2A4	61	+	+	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	+	+	+	-	-	-	-	-	+	+	-
2A5	62	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2A6	63	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2B2	64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
2B3	65	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
2C1	66	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2C2	67	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2C3	68	+	+	-	+	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	-	+	-	-	-	-	-	-	+	+	-
2C4	69	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2C6	70	-	-	+	-	-	-	-	+	-	+	-	+	+	-	+	-	-	-	+	-	-	+	+	-	-	-	-	-	+	+	-	+	-	-	-	-	+
2D2	71	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2D3	72	+	+	-	+	+	-	+	+	-	-	+	+	-	-	-	+	-	-	+	-	+	+	+	-	+	?	+	+	+	+	-	-	+	+	+	+	-
2D4	73	+	+	-	-	-	-	-	-	-	+	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	+	+	-	-	-	-	-	-	+	+	-
2E4	74	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-	+
2F2	75	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-
2F3	76	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
2F5	77	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2I2	78	+	-	-	-	-	-	-	-	-	+	-	+	+	-	+	+	-	-	+	-	+	+	-	-	-	-	-	+	-	+	-	-	-	-	?	-	-
2I3	79	+	-	-	-	+	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	+	+	-	-	-	-	-	+	-	+	-	-	-	-	+	-	-
2I4	80	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
2I5	81	-	-	-	+	-	-	-	-	-	+	-	+	+	-	+	+	-	-	+	-	-	?	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
2J2	82	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
2J5	83	-	-	-	+	-	-	+	-	-	+	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-
2L2	84	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	-	-	+	-	+	+	-	-	-	-	+	-	+	+	-	-	-	+	-	-	-
2L5	85	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-
2M1	86	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	-	-	-	-	-	+	-	-	+	+	+	-	+	-	-	-	+	-	-	+	+	-
2M6	87	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-
2N1	88	-	-	-	-	-	+	-	-	+	-	-	-	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2N2	89	-	-	-	-	-	+	-	-	+	-	-	-	+	+	-	-	+	-	-	+	-	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-
2N4	90	-	-	-	-	-	+	-	-	+	-	-	-	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2P1	91	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-
DOG	92	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DOG	92	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A23	93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A23	93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MQ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P(retained)	overall	0.198	0.154	0.165	0.176	0.176	0.209	0.154	0.167	0.187	0.484	0.154	0.253	0.308	0.154	0.451	0.297	0.220	0.110	0.198	0.198	0.189	0.256	0.242	0.169	0.198	0.133	0.297	0.143	0.143	0.244	0.143	0.143	0.154	0.143	0.189	0.167	0.176
	Average	0.204																																				

Table v

CELL LINE	HYBRID	1-09	1-11	1-20	1-27	1-28	1-30	1-32	1-42	1-46	1-49	1-58	1-60	K338	H32	H38	H162	H223	H254
A1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A3	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A6	3	-	+	+	-	-	+	-	+	+	-	-	-	-	+	-	-	-	-
B3	4	-	-	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	-
B4	5	-	+	+	-	-	+	+	+	-	+	-	-	-	-	-	+	+	-
B5	6	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-
C2	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
C3	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E2	9	+	-	-	-	-	-	-	+	-	+	?	-	-	+	-	+	-	-
E5	10	+	-	-	-	+	-	-	-	-	+	-	-	-	+	+	-	-	-
F1	11	+	-	-	-	-	-	-	+	-	-	-	-	-	+	+	+	+	-
G1	12	-	-	-	+	+	-	+	+	-	+	+	-	-	+	-	-	-	+
H6	13	?	+	-	?	-	-	+	+	-	-	-	-	-	-	-	-	-	+
I2	14	+	-	-	-	+	-	+	?	+	+	+	-	+	+	+	-	-	-
I4	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
I5	16	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-
I6	17	-	-	-	+	+	-	+	-	-	+	-	-	-	+	-	+	+	+
J3	18	-	+	-	-	+	-	-	-	+	-	-	+	+	+	+	-	-	-
K1	19	-	+	-	-	+	-	-	-	+	-	-	+	+	-	+	-	+	-
L3	20	-	+	-	-	+	-	-	-	+	+	+	-	+	+	+	-	+	-
L4	21	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-
M1	22	+	+	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-
M3	23	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
N1	24	-	-	-	-	-	-	-	-	-	?	-	-	-	-	-	-	-	-
N3	25	-	-	-	-	-	-	-	-	-	?	-	-	-	-	-	-	-	-
N5	26	+	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	+	+
N6	27	-	-	-	+	-	-	+	-	-	?	-	-	-	-	+	-	+	+
P1	28	+	-	-	-	-	-	-	+	-	-	+	-	-	-	+	-	-	-
P2	29	-	+	+	-	-	-	-	-	-	-	-	?	-	-	-	-	-	-
Q1	30	-	-	-	-	-	-	-	-	+	+	+	+	+	-	+	+	+	-
Q2	31	-	-	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-
Q4	32	-	-	-	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-
Q5	33	+	-	-	-	-	-	-	-	-	+	+	?	-	-	-	-	-	-
Q6	34	+	-	+	-	+	?	+	-	+	-	-	+	-	-	+	-	-	-
R1	35	?	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	+	+
R2	36	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
R3	37	+	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	+	+
S1	38	-	+	-	-	-	+	-	+	+	+	-	+	-	-	-	+	-	-
S2	39	-	-	+	+	-	+	+	-	-	+	+	-	-	+	-	+	+	+
S4	40	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
T5	41	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
T6	42	-	-	-	-	?	-	-	-	+	-	-	-	+	-	+	-	-	-
U2	43	+	-	-	-	-	-	-	-	+	-	-	-	+	-	+	-	-	-
V1	44	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-	-	-
V3	45	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
V4	46	-	+	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-
V5	47	+	-	+	+	-	+	+	+	+	+	-	+	+	+	+	+	+	+
W3	48	+	-	?	-	-	-	-	+	+	-	-	+	-	-	+	-	-	-
W4	49	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-
W5	50	-	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+
W6	51	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+
X1	52	+	-	-	-	-	-	-	+	+	-	-	+	-	-	+	-	-	-
X6	53	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Y2	54	-	-	-	+	-	-	+	-	+	-	-	+	+	-	+	-	-	+
Y3	55	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
Z1	56	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
Z3	57	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-
Z5	58	+	-	-	-	-	-	-	+	+	-	-	-	-	+	+	-	-	-
2A1	59	+	+	+	+	-	+	-	+	+	+	+	+	+	-	+	+	+	+
2A3	60	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2A4	61	-	-	?	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
2A5	62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2A6	63	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
2B2	64	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2B3	65	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2C1	66	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2C2	67	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-
2C3	68	-	+	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-
2C4	69	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-

Table v

CELL LINE	HYBRID	1-09	1-11	1-20	1-27	1-28	1-30	1-32	1-42	1-46	1-49	1-58	1-60	K336	H32	H38	H162	H223	H254
2C6	70	-	+	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-
2D2	71	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2D3	72	+	+	+	-	+	+	-	-	+	-	-	+	+	+	+	+	+	-
2D4	73	+	+	-	-	?	-	-	-	-	-	-	-	-	-	-	-	+	-
2E4	74	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2F2	75	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2F3	76	-	-	-	-	-		-	-	-	-		-	-	-	-	-	-	-
2F5	77	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2I2	78	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	+
2I3	79	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	+
2I4	80	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	+
2I5	81	+	-	?	+	+	-	+	-	-	-	-	-	-	-	-	-	-	+
2J2	82	-	-	-	-	-	-	?	-	+	+	-	-	+	-	+	+	+	-
2J5	83	+	-	-	-	-	-	+	-	-	+	+	-	-	+	-	-	-	-
2L2	84	-	-	-	+	-	-	+	+	+	-	-	-	-	-	+	-	-	+
2L5	85	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-
2M1	86	+	-	?	-	-	-	-	-	+	-	-	-	-	+	+	-	-	-
2M6	87	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
2N1	88	+	+	+	-	-	+	-	-	-	+	+	-	-	+	+	+	+	-
2N2	89	-	+	+	-	-	+	-	-	-	+	+	-	-	-	+	-	+	-
2N4	90	-	+	+	-	-	+	-	-	-	+	+	-	-	+	+	-	+	-
2P1	91	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	+	-
DOG	92	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DOG	92	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A23	93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A23	93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MQ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P(retained)	overall	0.348	0.231	0.161	0.189	0.227	0.133	0.267	0.211	0.286	0.250	0.167	0.135	0.154	0.198	0.451	0.165	0.231	0.198
	Average	0.222																	

References

Alderton D. (1994). Foxes Wolves and Wild Dogs of the World. Blandford, UK.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389 - 3402

The Kennel Club of Great Britain. (1998). The Kennel Club's Illustrated Breed Standards. Ebury Press, London, UK.

Anderson-Dear D. V. & Miller, J. R. (1994). Isolation of dinucleotide repeats from a pig chromosome 1-specific DNA library. *Mammalian Genome* **5**: 649 - 651.

Barendse, W., Armitage, S. M., Kossarek, L. M., Shalom, A., Kirkpatrick, B. W., Ryan, A.M., Clayton, D., Li, L., Neibergs, H. L., Zhang, N., *et al.* (1994). A genetic linkage map of the bovine genome. *Nature Genetics* **6**: 227 - 235

Barendse, W., Vaiman, D., Kemp, S. J., Sugimoto, Y., Armitage, S. M., Williams, J. L., Sun, H. S., Eggen, A., Agaba, M., Aleyasin, S. A., *et al.*, (1997). A medium-density genetic linkage map of the bovine genome. *Mammalian Genome* **8**: 21 – 28.

Beckman, J. S. & Weber, J. L. (1992). Survey of human and rat microsatellites. *Genomics* **12**: 627 - 631

Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouellette, B. F., Rapp, B. A. & Wheeler, D. L. (1999). GenBank. *Nucleic Acids Research* **27**: 12 - 17

Bentolila, S., Bach, J. M., Kessler, J. L., Bordelais, I., Cruaud, C., Weissenbach, J. & Panthier, J. J. (1999). Analysis of major repetitive DNA sequences in the dog (*Canis familiaris*) genome. *Mammalian Genome* **10**: 699 - 705.

Bicknell, D. C., Markie, D., Spurr, N. K. & Bodmer, W. F. (1991). The human chromosome content in human x rodent somatic cell hybrids analysed by a screening technique using *Alu* PCR. *Genomics* **10**: 186 - 192

Birren, B., Green, E. D., Hieter, P., Klapholz, S., Myers, R. M., Riethman, H. & Roskams, J. (Eds.) (1999). Genome Analysis: a laboratory manual. Volume 4. Mapping Genomes. Cold Spring Harbor Press, New York, USA.

Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sundem, S. L.F., Hawkins, G. A., Solinas-Toldo, S., Fries, R., Grosz, M. D., Yoo, J. & Beattie, C. W. (1994). A genetic linkage map for cattle. *Genetics* **136**: 619 – 639.

Boehnke M, Lange, K. & Cox, D. R. (1991). Statistical methods for multipoint radiation hybrid mapping. *American Journal of Human Genetics.* **49**: 1174 – 1188.

Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314 – 331.

Breen, M., Lindgren, G., Binns, M. M., Norman, J., Irvin, Z., Bell, K., Sandberg, K. & Ellegren, H. (1997). Genetical and physical assignments of equine microsatellites–first integration of anchored markers in horse genome mapping. *Mammalian Genome* **8**: 267 - 273.

Breen, M., Reimann, N., Bosma, A. A., Landon, D., Zijlstra, S., Bartnitzke, S., Switonski, M., Long, S. E., de Haan, A., Binns, M. M., Bullerdiek, J. & Langford, C. F. (1998). Standardisation of the chromosome nos. 22-38 of the dog (*Canis familiaris*) with the use of chromosome painting probes. *Proceedings of the 13th European Colloquium on Cytogenetics of Domestic Animals.* Budapest.

Breen, M., Langford, C. F., Carter, N. P., Holmes, N. G., Dickens, H. F., Thomas, R., Suter, N., Ryder, E. J., Pope, M., & Binns, M. M. (1999a). FISH mapping and identification of canine chromosomes. *Journal of Heredity* **90**: 27 - 30

Breen, M., Bullerdiek, J and Langford, C. (1999b). The DAPI banded karyotype of the domestic dog (*Canis familiaris*) generated using chromosome specific paint probes. *Chromosome Research* **7**: 401 - 406 (*et erratum* **7**: 575).

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *American Journal of Human Genetics* **63**: 861 - 869.

Brookes, A. J. (1999). The essence of SNPs. *Gene* **234**: 177 - 186

Bull, P. C., Thomas, G. R., Rommens, J. M., Forbes, J. R. & Cox, D. W. (1993). The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nature Genetics* **5**: 327 - 337

Burgoyne, P. S. (1982). Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Human Genetics* **61**: 85-90

Burmeister, M., Kim, S., Price, E. R., de Lange, T., Tantravahi, U., Myers, R. M. & Cox, D. R. (1991). A map of the distal region of the long arm of human chromosome 21 constructed by radiation hybrid mapping and pulsed-field gel electrophoresis. *Genomics* **9**: 19 - 30

Caetano-Anolles, G. (1996). Scanning of nucleic acids by in vitro amplification: new developments and applications. *Nature Biotechnology* **14**: 1668 - 1674.

Carter N. P. (1994). Bivariate chromosome analysis using a commercial flow cytometer. *Methods in Molecular Biology* **29**: 187 – 204.

Carrel, L., Cottle, A. A., Goglin, K. C. & Willard, H. F. (1999). A first-generation X-inactivation profile of the human X chromosome. *Proceedings of the National Academy of Sciences, USA* **96**: 14440 - 14444.

Ceccherini, I., Romeo, G., Lawrence, S., Breuning, M. H., Harris, P. C., Himmelbauer, H., Frischauf, A. M., Sutherland, G. R., Germino, G. G., Reeders, S. T. & Morton, N. E. (1992). Construction of a map of chromosome 16 by using radiation hybrids. *Proceedings of the National Academy of Sciences, USA* **89**: 104 – 108

Chang, K. S., Vyas, R. C., Deaven, L. L., Trujillo, J. M., Stass, S. A. & Hittelman, W. N. (1992) PCR amplification of chromosome-specific DNA isolated from flow cytometry-sorted chromosomes. *Genomics* **12**; (2): 307 - 312.

Cherif, D., Julier, C., Delattre, O., Derre, J., Lathrop, G. M. & Berger, R. (1990). Simultaneous localization of cosmids and chromosome R-banding by fluorescence microscopy: application to regional mapping of human chromosome 11. *Proceedings of the National Academy of Sciences, USA* **87**: 6639 - 6643.

Church, G. M. & Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences, USA*. **81**: 1991 – 1995

Cox, D. R., Burmeister, M., Price, E. R., Kim, S. & Myers, R. M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**: 245 - 250

Crawford, A. M., Dodds, K. G., Ede, A. J., Pierson, C. A., Montgomery, G. W., Garmonsway, H. G., Beattie, A. E., Davies, K., Maddox, J. F., Kappes, S. W., Stone, R. T., Nguyen, T. C., Penty, J. M., Lord, E. A., Broom, J. E., Buitkamp, J., Schwaiger, W., Epplen, J. T., Matthew, P., Matthews, M. E., Hulme, D. J., Beh, K. J., McGraw, R. A. & Beattie, C. W. (1995). An autosomal genetic linkage map of the sheep genome. *Genetics* **140**: 703 – 724.

Cremer, T., Lichter, P., Borden, J., Ward, D. C. & Manuelidis, L. (1988). Detection of chromosome aberrations in metaphase and interphase tumour cells by *in situ* hybridisation using chromosome-specific library probes. *Human Genetics*. **80**: 235 – 246.

Csink, A. K. & Henikoff, S. (1998). Something from nothing: the evolution and utility of satellite repeats *Trends in Genetics* **14**: 200 - 204

Dear, P. H. & Cook, P. R. (1993). Happy mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Research* **21**: 13 - 20.

Dear, P. H., Bankier, A. T. & Piper, M. B. (1998). A high-resolution metric HAPPY map of human chromosome 14. *Genomics* **48**: 232 - 241.

D'Esposito, M., Ciccodiola, A., Gianfrancesco, F., Esposito, T., Flagiello, L., Mazzarella, R., Schlessinger, D. & D'Urso, M. (1996). A synaptobrevin like gene in Xq28 pseudoautosomal region undergoes X inactivation. *Nature Genetics* **13**: 227 – 229

de Gortari, M. J., Freking, B. A., Cuthbertson, R. P., Kappes, S. M., Keele, J. W., Stone, R. T., Leymaster, K. A., Dodds, K. G., Crawford, A. M. & Beattie, C. W. (1998). A second-generation linkage map of the sheep genome. *Mammalian Genome* **9**: 204 - 209

Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matise, T. C., McKusick, K. B., Beckmann, J. S., Bentolila, S., Bihoreau, M., Birren, B., Browne, J., Butler, A., Castle, A. B., Chiannilkulchai, N., Clee, C., Day, P., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C., Fox, S., Gelling, S., Green, L., Harrison, P., Hocking, R., Holloway, E., Hunt, S., Keil, S., Lijnzaad, P., Louis-Dit-Sully, C., Ma, J., Mendis, A., Miller, J., Morissette, J., Muselet, D., Nusbaum, H., Peck, A., Rozen, S., Simon, D., Slonim, D., Staples, R., Stein, L., Stewart, E., Suchard, M., Thangarajah, T., Vega-Czarny, N., Webber, C., Wu, X., Hudson, J., Auffray, C., Nomura, N., Sikela, J., Polymeropoulos, M., James, M., Lander, E., Hudson, T., Myers, R., Cox, D., Weissenbach, J & Boguski, M. (1998). A physical map of 30,000 human genes. *Science* **282**: 744 - 746

Deschênes, S. M., Puck, J. M., Dutra, A. S., Somberg, R. L., Felsburg P. J. & Henthorn P. S. (1994). Comparative map of canine and human proximal Xq and genetic analysis of canine X-linked severe combined immunodeficiency. *Genomics* **23**: 62 – 68.

Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. & Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152 – 154.

Dickens, H. F., Holmes, N. G., Ryder, E. J., Thomas, R., Suter, N., Sampson, J., Langford, C. F., Ross, M., Carter, N. P., & Binns, M. M. (1999). Use of cosmid-derived and chromosome-specific microsatellites. *Journal of Heredity* **90**: 52 - 54

Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M. J., Ingalls, K. A., O'Connor, T. J., *et al.* (1996). A comprehensive genetic map of the mouse genome. *Nature* **380:** 149 – 152.

Disteche, C. M., Brannan, C. I., Larsen, A., Adler, D. A., Schorderet, D. F., Gearing, D., Copeland, N. G., Jenkins, N. A. & Park, L. S. (1992). The human pseudoautosomal GM-CSF receptor α subunit is autosomal in mouse. *Nature Genetics* **1**: 333 – 336.

Don R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991). "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**: 4008.

Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., Ainscough, R., Almeida, J. P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K. N., Beasley, O., Bird, C. P., Blakey, S., Bridgeman, A. M., Buck, D., Burgess, J., Burrill, W. D., O'Brien, K. P., *et al.* (1999). The DNA sequence of human chromosome 22. *Nature* **402**: 489 - 495.

Dutra, A. S., Mignot, E. & Puck, J. M. (1996). Gene localization and syntenic mapping by FISH in the dog. *Cytogenetics and Cell Genetics* **74**: 113 - 117.

Faraco, J., Lin, X., Li, R., Hinton, L., Lin, L. & Mignot, E. (1999). Genetic studies in narcolepsy, a disorder affecting REM sleep. *Journal of Heredity* **90**: 129 - 132

Fischer, A., Carmichael, K. P., Munnell, J. F., Jhabvala, P., Thompson, J. N., Matalon, R., Jezyk, P. F., Wang, P. & Giger, U. (1998). Sulfamidase deficiency in a family of Dachshunds: a canine model of mucopolysaccharidosis IIIA (Sanfilippo A). *Pediatric Research* **44**: 74 - 82

Fischer P. E., Spriggs, H. F., Holmes, N. G., Binns, M. M., Grace, C. & Nacheva, E. P. (1994). The karyotype of *Canis familiaris*. *Proceedings of the 11th European Colloquium on the Cytogenetics of Domestic Animals*: 30 - 32.

Fischer, P. E., Holmes, N. G., Dickens, H. F., Thomas, R., Binns, M. M., & Nacheva, E. P. (1996). The application of FISH techniques for physical mapping in the dog (*Canis familiaris*). *Mammalian Genome* **7**: 37 – 41.

Florijn, R. J., Bonden, L. A., Vrolijk, H., Wiegant, J., Vaandrager, J. W., Baas, F., den Dunnen, J. T., Tanke, H. J., van Ommen & G. J., Raap, A. K. (1995). High-resolution DNA Fiber-FISH

for genomic DNA mapping and colour bar coding of large genes. *Human Molecular Genetics* **4**: 831 - 836

Francisco, L. V., Langston, A. A., Mellersh, C. S., Neal, C. L. & Ostrander, E. A. (1996). A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian Genome* **7**: 359 – 362.

Fredholm, M. & Winterø A. K. (1995). Variation of short tandem repeats within and between species belonging to the Canidae family. *Mammalian Genome* **6**: 11 – 18.

Freije, D., Helms, C., Watson, M. S. & Donis-Keller, H. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**: 1784 – 1787.

Galloway, S. M., Hanrahan, V., Dodds, K., Potts, M. D., Crawford, A. M. & Hill, D. F. (1996). A linkage map of ovine chromosome X. *Genome Research* **6**: 667 – 677.

Gardner, D. G. (1996). Spontaneous squamous cell carcinomas of the oral region in domestic animals: a review and consideration of their relevance to human research. *Oral Diseases* **2**: 148 - 154

Geldwerth, D., Bishop, C., Guellaën, G., Koenig, M., Vergnaud, G., Mandel, J-L. & Weissenbach, J. (1985). Extensive DNA sequence homologies between the human Y chromosome and the long arm of the X chromosome. *European Molecular Biology Organisation Journal* **4**: 1739 – 1743

Gilbert, N. & Labuda, D. (1999). CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proceedings of the National Academy of Sciences, USA* **96**: 2869 - 2874.

Goodfellow, P. N. Banting, G., Sheer, D., Ropers, H-H., Caine, A., Ferguson-Smith, M. A., Povey, S. & Voss, R. (1983). Genetic evidence that a Y-linked gene in man is homologous to a gene on the X chromosome. *Nature* **302**: 346 – 349.

Gorski, J.L., Boehnke, M., Reyner, E. L. & Burright, E. N. (1992). A radiation hybrid map of the proximal short arm of the human X chromosome spanning Incontinentia Pigmenti 1 (IP1) translocation breakpoints. *Genomics* **14**: 657 - 665

Goss, S. & Harris, H. (1975). New method for mapping genes in human chromosomes. *Nature* **255**: 680 - 684.

Grandin T. (Ed.). (1998). Genetics and the behaviour of Domestic Animals. Academic Press,

Graphodatsky, A. S. in Halnan, C. R. E. (Ed.) (1989). Cytogenetics of Animals. CAB International, Oxford, UK.

Graves, J. A., Disteche, C. M. & Toder, R. (1998a). Gene dosage in the evolution and function of mammalian sex chromosomes. *Cytogenetics and Cell Genetics* **80**: 94 -103.

Graves, J. A. M., Wakefield, M. J. & Toder, R. 1998b The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Human Molecular Genetics* **7**: 1991 - 1996.

Green, P., Fall, K. & Crooks, S. (1990). Documentation for CRI-MAP, version 2.4, unpublished. URL: http://www.genome.iastate.edu/lab/CRIMAPwkshp/crimap-doc.html

Grimm, D. R., Goldman, T., Holley-Shanks, R., Buoen, L., Mendiola, J., School, L. B., Louis, C., Rohrer, G. A. & Lunney, J. K. (1997). Mapping of microsatellite markers developed from a flow-sorted swine chromosome 6 library. *Mammalian Genome* **8**: 193 – 199.

Guevara-Fujita, M. L., Loechel, R., Venta, P. J., Yuzbasiyan-Gurkan, V. & Brewer, G. J. (1996). Chromosomal assignment of seven genes on canine chromosomes by fluorescence *in situ* hybridization. *Mammalian Genome* **7**: 268 - 270.

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud-Homme, J-F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J. & Goodfellow, P. N. (1996). A radiation hybrid map of the human genome. *Human Molecular Genetics* **5**: 339 - 346

Halnan, C. R. E. (Ed.) (1989). Cytogenetics of Animals. CAB International, Oxford, UK.

He, L., Morris, S., Lennon, A., Clair, D. M. S., Porteous, D. J., Wright, A. F., Muir, W. J. & Blackwood, D. H. R. (1996). A genome-wide search for linkage in a large bipolar family: comparison of genotyping accuracy using di- and tetranucleotide repeat microsatellite markers. *Psychiatric Genetics* **6**: 123 – 129.

Henthorn, P. S., Somberg, R. L., Fimiani, V. M., Puck, J. M., Patterson, D. F. & Felsburg, P. J. (1994). IL-2R gamma gene microdeletion demonstrates that canine X-linked severe combined immunodeficiency is a homologue of the *Genomics* **23**: 69 - 74.

Herrtage, M. E., Seymour, C. A., White, R. A. S., Small, G. M. & Wight, D. G. D. (1987). Inherited copper toxicosis in the Bedlington terrier - the prevalence in asymptomatic dogs. *Journal of Small Animal Practice* **28**: 1141 – 1151.

Holmes, N. G., Mellersh, C. S., Humphreys, S. J., Binns, M. M., Holliman, A., Curtis, R. & Sampson, J. (1993a). Isolation and characterization of microsatellites from the canine genome. *Animal Genetics* **24**: 289 - 292.

Holmes, N. G., Humphreys, S. J., Binns, M. M., Curtis, R., Holliman, A. & Scott, A. M. (1993b). Characterization of canine microsatellites. *Experientia Supplementum* **67**: 415 - 420.

Holmes, N. G., Dickens, H. F., Parker, H. L., Binns, M. M., Mellersh, C. S., & Sampson, J. (1995). Eighteen canine microsatellites. *Animal Genetics* **26**: 132 – 133.

Holmes, N. G., Herrtage, M. E., Ryder, E. J. & Binns, M. M. (1998). DNA marker C04107 for copper toxicosis in a population of Bedlington terriers in the United Kingdom. *The Veterinary Record* **142**: 351 – 352.

Hu, Z., Rohrer, G., Murtaugh, M. P., Stone, R. T. & Beattie, C. W. (1997). Mapping genes to swine X chromosome provides reference loci for comparative mapping. *Mammalian Genome* **8**: 608 – 610.

Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J., Castle, A. B., Silva, J., Slonim, D. K., Baptista, R., Kruglyak, L., Xu, S-H. *et al.* (1995). An STS-based map of the human genome. *Science*. **270**: 1945 - 1954.

Hunt, P. A. & LeMaire, R. (1992). Sex-chromosome pairing: evidence that the behaviour of the pseudoautosomal region differs during male and female meiosis. *American Journal of Human Genetics* **50**: 1162 – 1170.

Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**: 971 - 983.

Jacob, H. J., Brown, D. M., Bunker, R. K., Daly, M. J., Dzau, V. J., Goodman, A., Koike, G., Kren, V., Kurtz, T., Lemmark, A., *et al.* (1995). A genetic linkage map of the laboratory rat, *Rattus norvegicus. Nature Genetics* **9**: 63 - 69

Jenkins, N. A. (1998). Plenary lecture at International Society for Animal Genetics meeting, 11/08/98, Auckland, New Zealand

Jiang, Z., Priat, C. & Galibert, F. (1998). Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. *Mammalian Genome* **9**: 577 – 587

Kappes, S. M., Keele, J. W., Stone, R. T., McGraw, R. A., Sonstegard, T. S., Smith, T. P., Lopez-Corrales, N. L. & Beattie, C. W. (1997). A second-generation linkage map of the bovine genome. *Genome Research* **7**: 235 - 249.

Kemp, S. J., Hishida, O., Wambugu, J., Rink, A., Longeri, M. L., Ma, R. Z., Da, Y., Lewin, H. A., Barendse, W. & Teale, A. J. (1995). A panel of polymorphic bovine, ovine and caprine microsatellite markers. *Animal Genetics* **26**: 299 - 306

Kermouni, A., Van Roost, E., Arden, K. C., Vermeesch, J. R., Weiss, S., Godelaine, D., Flint, J., Lurquin, C., Szikora, J. P., Higgs, D. R., Marynen, P. & Renauld, J. C. (1995). The IL9 receptor gene IL9R: genomic structure and chromosomal localisation in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter, 16pter and 18pter. *Genomics* **29**: 371 – 382.

Koenig, M., Moisan, J. P., Hellig, R. & Mandel, J. L. (1985). Homologies between X ands Y chromosomes detected by DNA probes: Localisation and evolution. *Nucleic Acids Research* **13**: 5285 – 5501

Kondo, Y., Mori, M., Kuramoto, T., Yamada, J., Beckmann, J., Simon-Chazottes, D., Montagutelli, X., Guénet, J-L. & Serikawa, T. (1993). DNA segments mapped by reciprocal use of microsatellite primers between mouse and rat. *Mammalian Genome* **4**: 541 – 576.

Konfortov, B. A., Cohen, H. M., Bankier, A. T. & Dear, P. H. A high resolution HAPPY map of *Disctyostelium discoideum* chromosome 6. (Submitted to *Genome Research*).

Korenberg, J. R., Yang-Feng, T., Schreck, R. & Chen, X. N. (1992). Using fluorescence *in situ* hybridization (FISH) in genome mapping. *Trends in Biotechnology* **10**: 27 - 32

Kuiper, M. T. (1998). Building a high-density genetic map using the AFLP technology. *Methods in Molecular Biology* **82**: 157 - 171

Kvaløy, K., Galvagni, F. & Brown, W. R. A. (1994). The sequence organisation of the long arm pseudoautosomal region of the human sex chromosomes. *Human Molecular Genetics* **3**: 771 – 778.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**: 77 - 79

Lander E. S. & Green P. (1987). Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences, USA* **84**: 2363 - 2367.

Lange K, Boehnke M., Cox, D. R. & Lunetta, K. L. (1995).. Statistical methods for polyploid radiation hybrid mapping. *Genome Research* **5**: 136 - 150.

Langford, C. F., Telenius, H., Carter, N. P., Miller, N. G. A. & Tucker, E. M. (1992). Chromosome painting using chromosome-specific probes from flow-sorted pig chromosomes. *Cytogenetics and Cell Genetics* **61**: 221 - 223.

Langford, C. F., Telenius, H., Miller, N. G. A., Thomsen, P. D. & Tucker, E. M. (1993). Preparation of chromosome-specific paints and the complete assignment of chromosomes in the pig flow karyotype. *Animal Genetics* **24**: 261 – 267.

Langford, C. F., Fischer, P. E., Binns, M. M. Holmes, N. G. & Carter, N. P. (1996). Chromosome-specific paints from a high-resolution flow karyotype of the dog. *Chromosome Research* **4**: 115 - 123.

Lathrop, G. M., Lalouel, J. M., Juliel C., *et al.* (1984).. Strategies for multilocus linkage analysis in humans. *Proceedings of the National Academy of Sciences, USA* **81**: 3443 - 3446.

Li, R., Mignot, E., Faraco, J., Kadotani, H., Cantanese, J., Zhao, B., Lin, X., Hinton, L., Ostrander, E. A., Patterson, D. F. & de Jong, P. J. (1999) Construction and characterisation of an eightfold redundant dog genomic bacterial artificial chromosome library. *Genomics* **58**: 9 -17

Lichter, P. Cremer, T., Borden J., Maneulidis, L. & Ward, D. C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridisation using recombinant DNA libraries. *Human Genetics* **80**: 224 – 234.

Lin, L., Faraco, J., Li, R., Kadotani, H., Rogers, W., Lin, X., Qiu, X., de Jong, P., Nishino, S. & Mignot, E. (1999) The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **8**: 365 - 376

Lingaas F., Sørenson A., Juneja R. K., Johannson S., Fredholm M., Winterø A. K., Sampson, J., Mellersh C. S., Curzon A., Holmes N. G., Binns M. M., Dickens H. F., Ryder E. J., Gerlach J., Bäumle E., Dolf G. (1997). Towards construction of a canine linkage map: establishment of 16 linkage groups. *Mammalian Genome* **8**: 218-221.

Liyanage, M., Coleman, A., du Manoir, S., Veldman, T., McCormack, S., Dickson, R. B., Barlow, C., Wynshaw-Boris, A., Janz, S., Wienberg, J., Ferguson-Smith, M. A., Schröck, E. & Ried, T. (1996). Multicolour spectral karyotyping of mouse chromosomes. *Nature Genetics* **14**: 312 - 315

Longmire, J. L., Brown, N. C., Meincke, L. J., Campbell, M. L., Albright, K. L., Fawcett, J. J., Campbell, E.W., Moyzis, R. K., Hildebrand, C. E., Evans, G. A., *et al.* (1993). Construction and characterization of partial digest DNA libraries made from flow-sorted human chromosome 16. *Genetic Analysis, Techniques and Applications* **10**: 69 - 76.

Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. (1990). Towards construction of high resolution map of the mouse genome using PCR analysed microsatellites. *Nucleic Acids Research* **18**: 4123 – 4129.

Lunetta, K. L. Boehnke, M., Lange, K. & Cox, D. R. (1996). Selected locus and multiple panel models for radiation hybrid mapping. *American Journal of Human Genetics* **59**: 717 – 725.

Lyons, L. A., Laughlin, T. F., Copeland, N. G., Jenkins, N. A., Womack, J. E. & O'Brien, S. J. (1997). Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics* **15**: 47 - 56.

Matise T. C., Perlin, M. & Chakravarti, A. (1994). Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genetics* **6**: 384 - 390

Matthey R. (1954). Chromosomes et systematique des canidés. *Mammalia* **18**: 225 - 230.

Mayrand, P. E., Corcoran, K. P., Ziegle, J. S., Robertson, J. M., Hoff, L. B. & Kronick, M. N. (1992). The use of fluorescence detection and internal lane standards to size PCR products automatically. *Applied and Theoretical Electrophoresis* **3**: 1 - 11

McCarthy, L. C., Terrett, J., Davis, M. E., Knights, C. J., Smith, A. L., Critcher, R., Schmitt, K., Hudson, J., Spurr, N. K & Goodfellow, P. N. (1997). A First-generation Whole Genome-Radiation Hybrid Map spanning the Mouse Genome. *Genome Research* **7**: 1153 – 1154.

McCarthy, L. & Soderlund C. (1998). Chapter 6 in Guide to Human Genome Computing, 2nd Edition. Academic Press, London, UK.

Mellersh C. S., Langston, A. A., Acland, G. M., Fleming, M. A., Ray, K. (1997). A linkage map of the canine genome. *Genomics* **46**: 326 – 336.

Mellersh, C. S., Hitte, C., Richman, M., Vignaux, F., Priat, C., Jouquand, S., Werner, P., Andre, C., DeRose, S., Patterson, D. F., Ostrander, E. A. & Galibert, F. (2000). An Integrated Linkage-Radiation Hybrid Map of the Canine Genome. *Mammalian Genome* **11**: 120 - 130

Menotti-Raymond, M., David, V. A., Lyons, L. A., Schäffer, A. A., Tomlin, J. F., Hutton, M. K. & O'Brien, S. J. (1999). A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* **57**: 9 - 23

Milatovich, A., Kitamura, T., Miyaiima, A. & Francke, U. (1993). Gene for the α subunit of the human interleukin-3 receptor (IL3RA) localised to the X-Y pseudoautosomal region. *American Journal of Human Genetics* **53**: 1146 – 1153.

Miller, J. R., Dixon, S. C., Miller, N. G. A., Tucker, E. M., Hindkjær, J. & Thomsen, P. D. (1992). A chromosome 1-specific DNA library from the domestic pig (*Sus scrofa domestica*). *Cytogenetics and Cell Genetics* **61**: 128 –131.

Miller, J. R. (1994). Use of porcine interspersed repeat sequences in PCR-mediated genotyping. *Mammalian Genome* **5**: 629 – 632

Millwood, I. Y., Bihoreau, M-T., Gauguier, D., Hyne, G., Levy, E. R., Kreutz, R., Lathrop, G. M. & Monaco, A. P. (1997). A gene-based genetic linkage and comparative map of the rat X chromosome. *Genomics* **40**: 253 – 261.

Modi, F. S., Fanning, T. G., Wayne, R. K. & O'Brien, S. J. (1988). Chromosomal localisation of satellite DNA sequences among 22 species of felids and canids (Carnivora).. *Cytogenetics and Cell Genetics* **48**: 208 – 213.

Monaco, A. P. & Larin, Z. (1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends in Biotechnology* **12**: 280 - 286

Morton N. E., Lindsten, J., Iselius, L. & Yee, S. (1982). Data and theory for a revised chiasma map of man. *Human Genetics* **62**: 266 – 270.

Mumm, S., Molini, B., Terrell, J., Srivastava, A. & Schlessinger, D. (1997). Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 6-kb resolution. *Genome Research* **7**: 307 – 314.

Murphy, W. J., Menotti-Raymond, M., Lyons, L. A., Thompson, M. A. & O'Brien, S. J. (1999). Development of a feline whole genome radiation hybrid panel and comparative mapping of human chromosome 12 and 22 loci. *Genomics* **57**: 1- 18.

Nath, J. & Johnson, K. L. (1998). Fluorescence *in situ* hybridization (FISH): DNA probe production and hybridization criteria. *Biotechnic and Histochemistry* **73**: 6 - 22

Neff, M. W., Broman, K. W., Mellersh, C. S., Ray, K., Acland, G. M., Aguirre, G. D., Ziegle, J. S., Ostrander, E. A. & Rine, J. (1999). A second-generation linkage map of the domestic dog, *Canis familiaris*. *Genetics* **151**: 803 – 820.

Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-Solis, R., Webster, T. D., Ledbetter, D. H. & Caskey, C. T. (1989). Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proceedings of the National Academy of Sciences, USA* **86**: 6686 - 6690

Nicholas F. W., Brown, C. S. & Le Tissier, P. R. (1998). Online Mendelian Inheritance in Animals. URL: http://www.angis.su.oz.au/Databases/BIRX/omia/omia_form.html

Nijman, I. J., Vastenburg, J. B., Williams, J. L., Olsaker, I., Thieven, U. & Lenstra, J. A. (1998). Thirteen bovine microsatellite markers that are polymorphic in sheep. *Animal Genetics* **29**: 474 - 475

Nižetić, D., Zehetner, G., Monaco, A. P., Gellen, L., Young, B. D. & Lehrach, H. (1991). Construction, arraying and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proceedings of the National Academy of Sciences, USA* **88**: 3233 – 3237.

O' Brien, S. J., Womack, J. E., Lyons, L. A., Moore, K. J., Jenkins, N. A. & Copeland, N. G. (1993). Anchored reference loci for comparative genome mapping in mammals. *Nature Genetics* **3**: 103 – 112.

Ohno, S. (1967). Sex chromosomes and sex-linked genes. Springer-Verlag, Berlin.

Ohno, S. (1973). Ancient linkage groups and frozen accidents. *Nature* **244**: 259 – 262

Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J. JR., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. & Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in Spino-cerebellar ataxia type 1. *Nature Genetics* **4**: 221 - 226

Ostrander, E. A., Sprague JR, G. F. & Rine, J. (1993). Identification and characterisation of dinucleotide repeat (CA)n markers for genetic mapping in Dog. *Genomics* **16**: 207 – 213.

Ott, J. in Human Genetic Disease: a practical approach, ed. K.E. Davies, IRL Press, Oxford, 1986; pp19-32

Ott, J. (1991). Analysis of human genetic linkage. Revised Edition. The Johns Hopkins Press Ltd., London

Page, D. C., Harper, M. E., Love, J. & Botstein, D. (1984). Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* **311**: 119 – 123

Page, D. C., Bieker, K., Brown, L. G., Hinton, S., Leppert, M., Lalouel, J. M., Lathrop, M., Nystrom-Lahti, M., de la Chapelle, A. & White, R. (1987). Linkage, physical mapping, and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. *Genomics* **1**: 243 – 256.

Palmer, S., Perry, J. & Ashworth, A. (1995). A contravention of Ohno's law in mice. *Nature Genetics* **10**: 472 – 476.

Patterson, D. F., Aguirre, G. A., Fyfe, J. C., Giger, U., Green, P. I., Haskins, M. E., Jezyk, P. F. & Meyers-Wallen, V. N. (1989). Is this a genetic disease? Journal of Small Animal Practice 30: 127 - 139

Patterson, D. F. (1993). Understanding and controlling inherited diseases in dogs and cats. *Tijdschrift voor Viergeneeskunde* **118**, Supplement **1**: 23S - 27S.

Pinkel, D., Landegent, J., Collins, C., *et al.*, (1988). Fluorescence *in situ* hybridisation with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences, USA* **85**: 9138 – 9142.

Piper, M. B., Bankier, A. T. & Dear, P. H. (1998). A HAPPY map of *Cryptosporidium parvum*. *Genome Research* **8**: 1299 - 1307.

Piumi, F., Schibler, L., Vaiman, D., Oustry, A. & Cribiu, E. P. (1998). Comparative cytogenetic mapping reveals chromosome rearrangements between the X chromosomes of two closely related mammalian species (cattle and goats). *Cytogenetics and Cell Genetics* **81**: 36 – 41.

Priat, C., Hitte, C., Vignaux, F., Renier, C., Jian, Z., Jouquand, S., Ceron, A., Andre, C. & Galibert, F. (1998). A whole-genome radiation hybrid map of the dog genome. *Genomics* **53**: 361 378.

Rabbitts, P., Impey, H., Heppell-Parton, A., Langford, C. F., Tease, C., Lowe, N., Bailey, D., Ferguson-Smith, M. & Carter, N. P. (1995). Chromosome specific paints from a high resolution karyotype of the mouse. *Nature Genetics* **9**: 369 – 375.

Rappold, G. A. (1993). The pseudoautosomal regions of the human sex chromosomes. *Human Genetics* **92**: 315 - 324.

Reimann N., Barnitzke, S. Bullerdiek, J., Schmitz, U., Rogalla, P., Nolte, I. & Rønne, M. (1996).. An extended nomenclature of the canine karyotype. *Cytogenetics and Cell Genetics.* **73**: 140 - 144.

Ried, T., Arnold, N., Ward, D. C. & Wienberg, J. (1993). Comparative high-resolution mapping of human and primate chromosomes by fluorescence *in situ* hybridization. *Genomics* **18**: 381 - 386.

Rohrer, G. A., Alexander, L. J., Keele, J. W., Smith, T. P & Beattie, C. W. (1994).. A microsatellite linkage map of the porcine genome. *Genetics* **136**: 231 - 245

Rønne, M., in Halnan, C. R. E. (Ed.) (1989). Cytogenetics of Animals. CAB International, Oxford, UK.

Ross, M. & Langford, C. F. (1997). in Genome Mapping, Dear, P. H. (Ed.) p172

Rothuizen, J., Wolfswinkel, J., Lenstra, J. A. & Frants, R. R. (1994). The incidence of mini- and micro-satellite repetitive DNA in the canine genome. *Theoretical & Applied Genetics* **89**: 403 - 406

Rugarli, E. I., Adler, D. A., Borsani, G., Tsuchiya, K., Franco, B., Hauge, X., Disteche, C., Chapmann, V. & Ballabio, A. (1995). Different chromosomal localisation of the Clcn4 gene in *Mus spretus* and C57BL/6J mice. *Nature Genetics* **10**: 466 - 471

Saiki, R. K., Bugawen, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* **324**: 163 - 166

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* **74**: 5463 - 5467

Schatzberg, S., Olby, N., Breen, M., Anderson, L., Langford, C., Dickens, H., Wilton, S., Zeiss, C., Binns, M., Kornegay, J., Morris, G., Sharp, N. (1999). Molecular analysis of a spontaneous dystrophin "knockout" dog. *Neuromuscular Disorders* **9**: 289 - 295.

Scherthan, H., Cremer, T., Arnason, U., Weier, H-U. Lima-de-Faria, A. & Fronicke, E. (1994). Comparative chromosome painting discloses homologous segments in distantly related mammals. *Nature Genetics* **6**: 342 – 347.

Schläpfer, J., Switonski, M., Schelling, C., Stranzinger, G., Gaillard, C., *et al.* (1998). The highly polymorphic canine microsatellite ZuBeCa2 is localised on canine chromosome 1q210 – q211. *Animal Genetics* **29**: 329 - 330

Schmitt K., Lazzeroni, L. C., Foote, S., Vollrath, D., Fisher, E. M. C., Goradia, T. M., Lange, K., Page, D. C. & Arnheim, N. (1994). Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing: do double crossovers occur during male meiosis? *American Journal of Human Genetics.* **55**: 423 – 430.

Schröck, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M. A., Ning, Y., Ledbetter, D. H., Bar-Am, I., Soenksen, D., Garini, Y. & Ried, T. (1996). Multicolour-spectral karyotyping of human chromosomes. *Science* **273**: 494 - 497.

Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T. J., *et al.* (1996). A gene map of the human genome. *Science* **274**, 540 - 546

Seabright, M. (1971). A rapid banding technique for human chromosomes. *Lancet* **ii**: 971 – 972.

Selden, J. R., Moorhead, P. S., Oehlert, M. L. & Patterson, D. F. (1975). The Giemsa banding pattern of the canine karyotype. *Cytogenetics and Cell Genetics* **15**: 380 - 387.

Sharp, N. J., Kornegay, J. N., Van Camp, S. D., Herbstreith, M. H., Secore, S. L., Kettle, S., Hung, W. Y., Constantinou, C. D., Dykstra, M. J., Roses, A. D., *et al.* (1992). An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. *Genomics* **13**: 115 - 121.

Sheldon J.W. (1992). Wild Dogs: The Natural History of the Nondomestic Canidae. Academic Press.

Shelton, G. D. (1999). Acquired myasthenia gravis: what we have learned from experimental and spontaneous animal models. *Veterinary Immunology and Immunopathology* **69**: 239 - 249

Shibuya, H., Collins, B. K., Huang, TH-M. & Johnson, G. S. (1994). A polymorphic $(AGGAAT)_n$ tandem repeat in an intron of the canine von Willebrand factor gene. *Animal Genetics* **25**: 122

Slonim, D., Kruglyak, L., Stein, L. & Lander, E. (1997). Building human genome maps with radiation hybrids. *Journal of Computational Biology* **4**: 487 – 504.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* **321**: 674 -679.

Speicher, M. R., Gwyn Ballard, S. & Ward, D. C. (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genetics* **12**: 368 - 375.

Solinas-Toldo, S., Mezzelani, A., Hawkins, G. A., Bishop, M. D., Olsaker, I., Mackinlay, A., Ferretti, L. & Fries, R. (1995). Combined Q-banding and fluorescence *in situ* hybridization for the identification of bovine chromosomes 1 to 7. *Cytogenetics and Cell Genetics* **69**: 1 - 6.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E. & Moyzis, R. K. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* **10**: 807 – 815.

Stallings, R. L. (1995). Conservation and evolution of (CT)n/(GT)n microsatellite sequences at orthologous positions in diverse mammalian genomes. *Genomics* **25**: 107 - 113

Stein, L. (1996). RHMAPPER, Installation and user's guide. Unpublished. URL: http://www-genome.wi.mit.edu/ftp/pub/software/rhmapper.

Stewart, A., McKusick, K. B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, Harris, M., Hussain, S., Lee, R., Maratukulum, A., O'Connor, K., Perkins, S., Piercy, M., Qin, F., Reif, T., Sanders, C., She, X., Sun, W., Tabar, P., Voyticky, S., Cowles, S., Fan, J., Mader, C., Quackenbush, J., Myers, R. & Cox, D. (1997). An STS-Based Radiation Hybrid Map of the Human Genome. *Genome Research* **7**: 422 - 433

Stone, D. M., Jacky, P. B. & Prieur, D. J. (1991). The Giemsa banding pattern of canine chromosomes, using a cell synchronisation technique. *Genome* **34**: 407 - 412.

Sturtevant, A. H., (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology* **14**: 43 - 59.

Suter, N. (1998). The isolation and characterisation of cosmid-derived canine microsatellite DNA sequences. Ph.D. dissertation, University of Leicester UK.

Switonski, M., Reimann, N., Bosma, A. A., Long, S., Barnitzke, S. *et al.* (1996). Report on the progress of standardisation of the G-banded canine (*Canis familiaris*) karyotype. *Chromosome Research* **4**: 306 - 309

Tagle, D. A. & Collins, F. S. (1992). An optimised *Alu*-PCR primer pair for human-specific amplification of YACs and somatic cell hybrids. *Human Molecular Genetics* **1**: 121 – 122.

Tanzi, R. E., Petrukhin, K., Chernov, I., Pellequer, J. L., Wasco, W., Ross, B., Romano, D. M., Parano, E., Pavone, L., Brzustowicz, L. M., *et al.* (1993). The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. *Nature Genetics* **5**: 344 - 350

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463 – 6471

Tautz, D. & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**: 4127 – 4138

Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjöld, M, Ponder, B. A. J. & Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**: 718 – 725.

Terwilliger, J. & Ott, J. (1994). Handbook for Human Genetic Linkage. Johns Hopkins University Press, Baltimore, USA.

Thomas, R., Breen, M., Langford, C.F. & Binns, M. M. (1999). Zoo-FISH analysis of dog chromosome 5: identification of conserved synteny with human and cat chromosomes. *Cytogenetics and Cell Genetics* **87**: 4 - 10.

Thompson E.A. (1987). Crossover counts and likelihood in multipoint linkage analysis. *IMA Journal of Mathematics Applied in Medicine and Biology* **4**: 93 - 108.

Toder R., Gläser, B., Schiebel, K., Wilcox, S. A., Rappold, G., Graves, J. A. M. & Schempp, W. (1997). Genes located in and near the human pseudoautosomal region are located in the X-Y pairing region in dog and sheep. *Chromosome Research* **5**: 301 - 306

Trask B. J. Chapter 7 in Birren, B., Green, E. D., Hieter, P., Klapholz, S., Myers, R. M., Riethman, H. & Roskams, J. (Vol. Eds.) (1999). Genome Analysis: a laboratory manual. Volume 4. Mapping Genomes. Cold Spring Harbor Press, New York, USA.

Trask, B. J. (1991). Fluorescence *in situ* hybridisation: applications in cytogenetics and gene mapping. *Trends in Genetics* **7**: 149 - 154

Twedt, D. C., Sternlieb, I. & Gilbertson, S. R. (1979). Clinical, morphologic, and chemical studies on copper toxicosis of Bedlington Terriers. *Journal of the American Veterinary Medical Association* **175**: 269 - 275

Vaiman, D., Schibler, L., Bourgeois, F., Oustry, A., Amigues, Y. & Cribiu, E. P. (1996). A genetic linkage map of the male goat genome. *Genetics* **144**: 279 – 305.

van de Sluis, B. J. A., Breen, M., Nanji, M., van Wolferen, M., de Jong, P., Binns, M. M., Pearson, P. L., Kuipers, J., Rothuizen, J., Cox, D. W., Wijmenga, C. & van Oost, B. A. (1999). Genetic mapping of the copper toxicosis locus in Bedlington terriers to dog chromosome 10, in a region syntenic to human chromosome region 2p13-p16. *Human Molecular Genetics* **8**: 501 – 507.

VanDevanter, D. R., Choongkittawom, N. M., Dyer, K. A., Aten, J., Otto, P., Behler, C., Bryant, E. M. & Rabinovitch, P. S. (1994) Pure chromosome-specific PCR libraries from single sorted chromosomes. *Proceedings of the National Academy of Sciences, USA* **91**; (13): 5858 - 5862.

Venta, P. J. Brouillette, J. A., Yuzbasiyan-Gurkan, V. & Brewer, G. J. (1996). Gene-specific universal mammalian sequence-tagged sites: application to the canine genome. *Biochemical Genetics* **34**: 321 - 341

Vilà, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandall, K. A., Lundeberg, J. & Wayne, R. K. (1997). Multiple and ancient origins of the domestic dog. *Science* **276**: 1687 - 1689

Vooijs, M., Yu, L. C., Tkachuk, D., Pinkel, D., Johnson D, & Gray, J. W. (1993) Libraries for each human chromosome, constructed from sorter-enriched chromosomes by using linker-adaptor PCR. *American Journal of Human Genetics* 52; (3): 586 - 597.

Walter, M. A., Spillett, D. J., Thomas, P., Weissenbach, J. & Goodfellow, P. N. (1994). A method for constructing radiation hybrid maps of whole genomes. *Nature Genetics* **7**: 22 - 28.

Watanabe, T. K., Bihoreau, M. T., McCarthy, L. C., Kiguwa, S. L., Hishigaki, H., Tsuji, A., Browne, J., Yamasaki, Y., Mizoguchi-Miyakita, A., Oga, K., Ono, T., Okuno, S., Kanemoto, N., Takahashi, E., Tomita, K., Hayashi, H., Adachi, M., Webber, C., Davis, M., Kiel, S., Knights, C., Smith, A., Critcher, R., Miller, J., James, M. R., *et al.* (1999). A radiation hybrid map of the rat genome containing 5,255 markers. *Nature Genetics* **22:** 27 - 36

Waters, D. J. & Bostwick, D. G. (1997). The canine prostate is a spontaneous model of intraepithelial neoplasia and prostate cancer progression. *Anticancer Research* **17**: 1467 - 1470

Wayne, R. K. (1993). Molecular Evolution of the dog family. *Trends in Genetics* **9**: 218 – 224.

Weber, J. L. & May, P. E. (1989). Abundant class of human DNA polymorphism's which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388 – 396

Weber, J. L. (1990). Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**: 524 - 530

Weier, H. U., Polikoff, D., Fawcett, J. J., Greulich, K. M., Lee, K. H., Cram, S., Chapman, V. M. & Gray, J. W. (1994). Generation of five high-complexity painting probe libraries from flow-sorted mouse chromosomes. *Genomics* **21**: 641 - 644.

Werner, P., Raducha, M. G., Prociuk, U., Henthorn, P. S. & Patterson, D. F. (1997). Physical and linkage mapping of human chromosome 17 loci to dog chromosomes 9 and 5. *Genomics* **42**: 74 – 82.

Werner, P., Mellersh, C. S., Raducha, M. G., DeRose, S., Acland, G. M., Prociuk, U., Wiegand, N., Aguirre, G. D., Henthorn, P. S., Patterson, D. F. & Ostrander, E. A. (1999). Anchoring of canine linkage groups with chromosome-specific markers. *Mammalian Genome* **10**: 814 - 823.

Wienberg, J., Jauch, A., Stanyon, R.,& Cremer, T. (1990). Molecular cytotaxonomy of primates by chromosomal *in situ* suppression hybridisation. *Genomics* **8**: 347 – 350.

Wienberg, J., Stanyon, R., Jauch, A. & Cremer, T. (1992). Homologies in human and *Macaca fuscata* chromosomes revealed by *in situ* hybridisation with human chromosome-specific DNA libraries. *Chromosoma* **101**: 265 – 270.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531 - 6535.

Willard, H. F., Goss, S. J., Holmes, M. T. & Munroe, D. L. (1985). Regional localisation of the phosphoglycerate kinase gene and pseudogene on the human X chromosome and assignment of a related DNA sequence to chromosome 19. *Human Genetics* **71**: 138 - 143

Womack, J. E., Johnson, J. S., Owens, E. K., Rexroad III, C. E., Schläpfer, J. & Yang Y-P. (1997). A whole-genome radiation hybrid panel for bovine gene mapping. *Mammalian Genome* **8**: 854 - 856.

Xiao, C., Tsuchiya, K. & Sutou, S. (1998). Cloning and mapping of bovine ZFX gene to the long arm of the X-chromosome (Xq34) and homologous mapping of ZFY gene to the distal region of the short arm of the bovine (Yp13), ovine (Yp12-p13), and caprine (Yp12-p13) Y chromosome. *Mammalian Genome* **9**: 125 – 130

Yang, F., O'Brien, P. C., Milne, B. S., Graphodatsky, A. S., Solanky, N., Trifonov, V., Rens, W., Sargan, D. & Ferguson-Smith, M. A. (1999). A complete comparative chromosome map for

the dog, red fox, and human and its integration with canine genetic maps. *Genomics* **62**: 189 - 202.

Yerle, M, Pinton, P., Robic, A., Alfonso, A., Palvadeau, Y., Delcros, C., Hawken, R., Alexander, L., Beattie, C., Schook, L., Milan, D. & Gellin J. (1998). Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenetics and Cell Genetics* **82**: 182 – 188

Yuzbasiyan-Gurkan V., Halloran Blanton, S., Cao, Y., Ferguson, P., Li, J., Venta, P. & Brewer, G. J. (1997). Linkage of microsatellite marker to the canine copper toxicosis locus in Bedlington terriers. *American Journal of Veterinary Research* **58**: 23 – 27.

Yuzbasiyan-Gurkan, V., Wagnitz, S., Blanton, S. H. & Brewer, G. J. (1993). Linkage studies of the esterase D and retinoblastoma genes to canine copper toxicosis: a model for Wilson disease. *Genomics* **15**: 86 - 90

Zajc I., Mellersh, C. S. & Sampson, J. (1997). Variability of canine microsatellites within and between different dog breeds. *Mammalian Genome* **8**: 182 – 185.

Zheng, K., Thorner, P. S., Marrano, P., Baumal, R. & McInnes, R. R. (1994). Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked hereditary nephritis resulting from a single base mutation in the gene encoding the alpha 5 chain of collagen type IV. *Proceedings of the National Academy of Sciences, USA* **91**: 3989 - 3993.

Zischler, H., Nanda, I., Schafer, R., Schmid, M. & Epplen, J. T. (1989). Digoxigenated oligonucleotide probes specific for simple repeats in DNA fingerprinting and hybridization *in situ*. *Human Genetics* **82**: 227 - 233.

PERGAMON

Neuromuscular Disorders 9 (1999) 289–295

Molecular analysis of a spontaneous dystrophin 'knockout' dog

Scott J. Schatzberg[a,*], Natasha J. Olby[a], Matthew Breen[b], Louise V.B. Anderson[c], Cordelia F. Langford[d], Helen F. Dickens[b], Stephen D. Wilton[e], Caroline J. Zeiss[f], Matthew M. Binns[b], Joe N. Kornegay[g], Glenn E. Morris[h], Nicholas J.H. Sharp[a]

[a]*College of Veterinary Medicine, North Carolina State University, 4700 Hillsborough Street, Raleigh, NC 27606, USA*
[b]*Animal Health Trust, P.O. Box 5, Lanwades Park, Kentford Newmarket, Suffolk, CB8 7UU, UK*
[c]*Neurobiology Department, University Medical School, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK*
[d]*The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK*
[e]*Australian Neuromuscular Research Institute, Medical Center, Nedlands, Western Australia 6009, Australia*
[f]*Cornell University, Baker Institute, Hungerford Hill Road, Ithaca, New York 14853-6401, USA*
[g]*College of Veterinary Medicine, University of Missouri-Columbia, Clydesdale Hall, Columbia, MO 6821,1 USA*
[h]*MRIC Biochemistry Group, N.E. Wales Institute, Wrexham LL11 2AW, UK*

Received 20 July 1998; received in revised form 22 January 1999; accepted 26 January 1999

Abstract

We have determined the molecular basis for skeletal myopathy and dilated cardiomyopathy in two male German short-haired pointer (GSHP) littermates. Analysis of skeletal muscle demonstrated a complete absence of dystrophin on Western blot analysis. PCR analysis of genomic DNA revealed a deletion encompassing the entire dystrophin gene. Molecular cytogenetic analysis of lymphocytes from the dam and both dystrophic pups confirmed a visible deletion in the p21 region of the affected canine X chromosome. Utrophin is up-regulated in the skeletal muscle, but does not appear to ameliorate the dystrophic canine phenotype. This new canine model should further our understanding of the physiological and biochemical processes in Duchenne muscular dystrophy. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Dystrophin; Duchenne muscular dystrophy; Canine; Animal models; Fluorescence in situ hybridization; Polymerase chain reaction

1. Introduction

Duchenne muscular dystrophy (DMD) is the second most common human genetic disorder, affecting 1 in 3500 live-born males. One-third of DMD cases occur sporadically without any prior family history. Affected boys are crippled from progressive muscle wasting and weakness by 12 years of age, and generally die by the third decade from respiratory or cardiac failure. Becker muscular dystrophy (BMD), an allelic variant of DMD, is a less severe form of muscle weakness. Both DMD and BMD are X-linked disorders caused by mutations in the DMD gene. This gene codes for dystrophin, a key cytoskeletal protein that is thought to strengthen muscle fiber membranes.

The dystrophin gene spans over 2.4 megabases in the p21 region of the human X-chromosome [1]. The gene is regulated by at least eight different promoters in a cell-specific manner. The Purkinje (P) and cortical (C) promoters, which mainly express their dystrophin isoforms in brain, are also transcriptionally active at low levels in skeletal and cardiac muscle [2,3]. The muscle-specific or M promoter is responsible for the majority of full-length dystrophin expression in both skeletal and cardiac muscle [4]. Intragenic dystrophin deletions are responsible for approximately two-thirds of DMD and BMD cases [5]. Although a large percentage of these mutations can be detected by PCR screening and RFLP analysis [6–8], the majority of dystrophin deletions cannot be visualized using classical cytogenetics. The first patient (B.B.) identified with a cytogenetically visible deletion in Xp21 suffered from a total of four X-linked conditions [9]. Other patients with cytogenetically detectable X-chromosomal deletions have been identified, but such cases

* Corresponding author. E-mail: scott_schatzberg@ncsu.edu

0960-8966/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0960-8966(99)00011-5

are extremely unusual [10–12]. More recently, advances in fluorescence in situ hybridization (FISH) technology have allowed the detection of female carriers with major deletions in the dystrophin gene [13,14].

Several animal models for DMD have been characterized at a molecular level including the *mdx* mouse, the dystrophic cat, and two dystrophic dog breeds. These mutations include a premature stop codon in exon 23 of the *mdx* mouse [15], a deletion of the muscle and Purkinje dystrophin promoters in the dystrophic cat [16], a splice site mutation in the dystrophic golden retriever [17], and a premature stop codon in exon 58 of the dystrophic rottweiler [18]. To date, no animal model has been described with a cytogenetically detectable dystrophin deletion. We have recently identified two German short-haired pointer (GSHP) littermates with generalized muscle atrophy, dilated cardiomyopathy, elevated CK levels, and no dystrophin immunoreactivity (Olby et al., in preparation). Here we describe a major deletion in p21 region of the canine X chromosome, confirmed by molecular cytogenetic analysis to encompass the entire dystrophin gene.

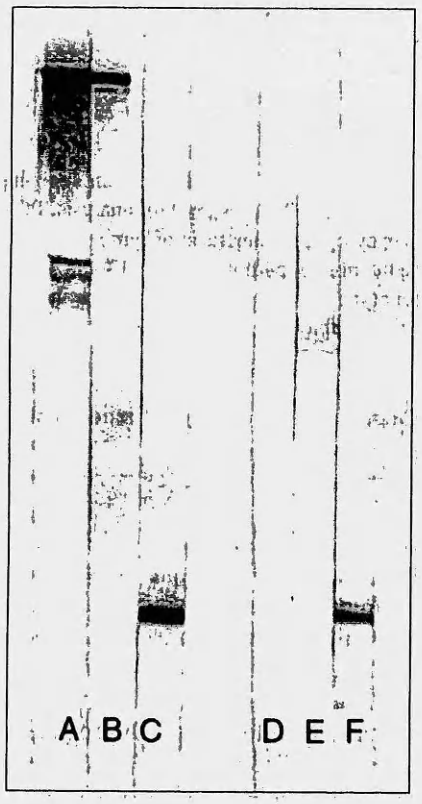


Fig. 1. Western blot analysis using monoclonal antibodies directed against dystrophin's rod domain (Dy4/6D3) and carboxy-terminus (Dy8/6C5) and against the dystrophin associated glycoprotein, β-dystroglycan (43DAG/8D5). Normal canine skeletal muscle shows the expected 427 kDa dystrophin band when labeled with the rod (lane A) and C-terminal (lane B) anti-dystrophin antibodies. Dystrophic GSHP skeletal muscle displays no immunoreactivity with either the rod (lane D) or C-terminal (lane E) antibodies. Both the normal dog (lane C) and affected GSHP (lane E) display the 43 kDa β-dystroglycan band, although the levels were slightly reduced in the GSHP.

2. Materials and methods

2.1. Animals

Two GSHP littermates (dogs I and II) born to a healthy dam presented at 5 months of age for exercise intolerance. Clinical details of the affected dogs will be described elsewhere (Olby et al., in preparation) and are briefly summarized in the results. The diagnosis of skeletal myopathy was established by elevated serum creatine kinase levels and from histopathologic examination of a triceps muscle biopsy taken from one of the GSHPs (dog II).

2.2. Western blot analysis

Dystrophin Western blot analysis, using both the rod and C-terminal anti-dystrophin antibodies (Dy4/6D3 and Dy8/6C5), was performed on GSHP dog II and on normal canine skeletal muscle as a positive control (Fig. 1). The methods have all been described previously [19,20]. The blots were also labeled with an antibody (43DAG/8D5) to β-dystroglycan, as a representative member of the dystrophin associated glycoprotein complex (Fig. 1). Blots of skeletal muscle from GSHP dog II were also labeled with the utrophin antibody, MANCHO3 (Fig. 2) [21]. Utrophin levels from one normal dog and from five dystrophin-deficient GRMD dogs, which over-express utrophin, were used for comparison. The five GRMD dogs used for comparison varied in clinical phenotype. Three of these GRMD dogs were severely affected (Fig. 2; lanes 2, 4, and 7) and died by 6 months of age; two of these dogs, which were phenotypically similar to GSHP dog II, were mildly affected (Fig. 2; lanes 9 and 11) and lived 6 and 8 years, respectively.

2.3. Genomic DNA extraction

Blood was collected in EDTA tubes from one normal dog and from the two GSHPs with muscular dystrophy. Genomic DNA was extracted from 300 μl aliquots of blood using the QIAamp blood & tissue kit (Qiagen, Chatsworth, CA).

2.4. Oligonucleotide primer design

Oligonucleotide PCR primers were designed based on sequences of the canine dystrophin M promoter [22], canine dystrophin exons 5 and 79 (Wilton et al., in preparation), canine tissue inhibitor of metaloproteinase-1 (TIMP-1) intron 2 (Zeiss et al., in preparation), and canine retinitis pigmentosa GTP-ase regulator (RPGR) introns 4 and 9 [23]. All oligonucleotide sequences that have not been published previously are indicated in Table 1. Primers were used for PCR after fractionation on a 20% polyacrylamide gel.

2.5. Polymerase chain reaction

Amplifications for the dystrophin M promoter and exon

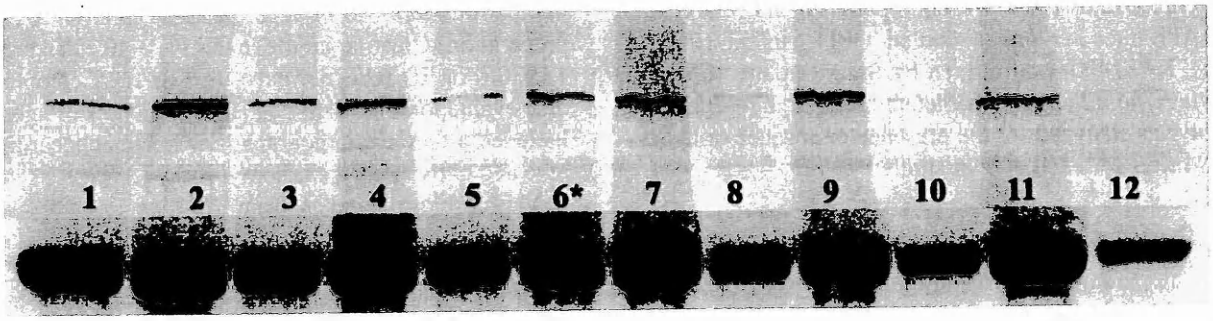


Fig. 2. Western blot analysis using a monoclonal antibody (MANCH03) directed against canine utrophin (top). Lanes 1, 3, 5, 8, 10 and 12 show faint utrophin bands in normal dog muscle (120, 100, 80, 60, 40 μg of muscle protein, respectively). The blot revealed clear bands of apparently identical intensity in the dystrophic GSHP (lane 6) skeletal muscle (120 μg), three severe GRMD dogs (lanes 2, 4, 7; 120 μg each), and two mild GRMD dogs (lanes 9 and 11; 120 μg). Myosin heavy chain bands (bottom) confirm the relative muscle protein abundance in each lane.

79, RPGR and TIMP-1 were optimized using normal canine genomic DNA and were then conducted on the DNA from both GSHPs. Amplification was carried out in 50 μl reactions containing PCR buffer diluted to 1X, 2 mM $MgCl_2$, 200 mM dNTP, 150 ng of each primer, 500 ng of genomic DNA as template and 0.25 units of Expand Taq Polymerase (Boerhinger Mannheim, Indianapolis, IN). After an initial denaturation at 99°C for 5 min, 35 cycles of 94°C for 1 min, 55°C for 30 s and 72°C for 1 min were carried out.

The integrity of genomic DNA was tested using primers that amplify the TIMP-1 gene and RPGR genes as positive controls. These amplifications were also used to approximate the proximal X-chromosomal deletion border. All PCR reactions were run concurrently with a water template as a negative control.

An 8 μl aliquot of each reaction was fractionated on a 2% agarose gel in TAE buffer. Agarose gels were stained with ethidium bromide and the resultant bands were visualized using UV light.

2.6. Chromosome preparation and FISH

Metaphase chromosomes of the dam and the two affected GSHP littermates (dogs I and II) were prepared by 72 h cultures of peripheral lymphocytes in the presence of pokeweed mitogen. Conventional harvesting procedures of colcemid and hypotonic treatments followed by fixation in (3:1) methanol:glacial acetic acid were used. Slides to be used for FISH analysis were passed through an ethanol series (70, 90 and 100%) and the chromosomes were then denatured in 70% formamide, 2× SSC for 2 min at 65°C, passed through the ethanol series a second time, and air dried before use.

Two plasmid clones (x71 and x74) from a flow sorted X-chromosome library (average insert size 3.5 kb), known to map to the p21 region of the canine X-chromosome, were labeled with digoxigenin-11-dUTP. Two additional clones, known to map distal (x9) and proximal (x13) to the p21 region, were labeled with biotin-16-dUTP. All labeling was performed by nick translation reactions in which the concentration of the DNAse-I was optimized to produce an average fragment size of 200–300 bp. Labeled DNAs (150 ng) were mixed with 15 μg of sonicated canine genomic DNA, precipitated, and resuspended in 15μl of a hybridization buffer comprising 50% formamide, 2 × SSC and 10% dextran sulphate. The DNA mixture was denatured at 70°C for 10 min and pre-annealed at 37°C for 30 min prior to being added onto the denatured chromosomes under a 22 × 22 mm coverslip and sealed with cowgum. The hybridization was continued for 16 h at 37°C in a humidified chamber. Post-hybridization stringency washes were as described previously [24]. The biotinylated probes were detected by Texas Red Avidin DCS (Vector) and the digoxigenin labeled probes were detected by FITC conjugated mouse anti-digoxin (Sigma). The chromosomes were counterstained in 4′-6′diaminidino-2phenylindole (DAPI) prior to being mounted with Vectashield (Vector) and sealed with a coverslip. Images were captured using a fluorescence

Table 1

Oligonucleotide primers used for PCR screening of canine dystrophin exon 79, tissue inhibitor of metaloproteinase-1 (TIMP-1) intron 2, and retinitis pigmentosa GTP-ase regulator (RPGR) intron 9 (1150 bp)

Primer	Sequence (5′–3′)	Primer	Sequence (5′–3′)	Product size (bp)
TIMP-F	GTC GGT CTG GTT GAC TTC TGC	TIMP-R	TCC TGC TGT TGC TGT GGC TGA C	851
RPGR-F	CAA TCA GTT CAT TCC CAC TCT GTG C	RPGR-R	CAT TTC TTC TGT CCC ACC AAG TCG	1150
DYS79-F (3′UTR)	CAA TGT AGG AAG CCT TTT CCA C	DYS79-R	AAC CAA AGT GAG GTA GAA ATA GC	2100

microscope (Axiophot, Zeiss) equipped with an FITC/Texas Red/DAPI filter set and a cooled CCD camera (Photometrics, KAF1400) both driven by SmartCapture software (Vysis). A minimum of 30 metaphases per hybridization reaction was analyzed for the presence or absence of hybridization sites for each animal studied.

3. Results

3.1. Patient evaluation

Two male GSHP littermates with skeletal myopathy were referred to NCSU CVM. One dog was initially presented at 5 months of age, and both dogs were evaluated at 9 months of age. Both dogs had marked generalized skeletal muscle atrophy with trismus. They had several skeletal abnormalities including brachygnathism, adduction of their hocks and concavity of their rib cage just proximal to the sternum. Both dogs exhibited the same abnormalities, but one dog (dog I) was much more severely affected than the other, with severe stunting of growth. Both dogs could only walk for short distances before tiring. No other abnormalities were noted on general physical or neurological examination. Chemistry panels revealed markedly elevated serum creatine kinase activity in both dogs (40 200 IU/l in dog I and 12 950 IU/l in dog II, normal: <160 IU/l). Cardiac evaluation including chest radiographs, electrocardiograms and echocardiography of both of these dogs at this time was unremarkable. The clinical progression of these dogs over a number of years will be reported elsewhere.(Olby et al., in preparation) Creatine kinase levels were 40 200 (dog I) and 13 930 (dog II) IU/l (normal <160 IU/l). A complete cardiac evaluation of GSHP dog II at 20 months of age, which included a thoracic radiograph, ECG, and echocardiogram, revealed early dilated cardiomyopathy [25].

Muscle biopsies taken from the triceps, vastus lateralis and cranial tibial muscles revealed severe changes consistent with a dystrophinopathy. Histopathologic features identified on hematoxylin and eosin stained sections included increased perimysial fibrosis, marked variation in fiber size and the presence of hyalinized fibers. There were also clusters of small, rounded, basophilic fibers with intracytoplasmic nuclei that represented regenerating fibers.

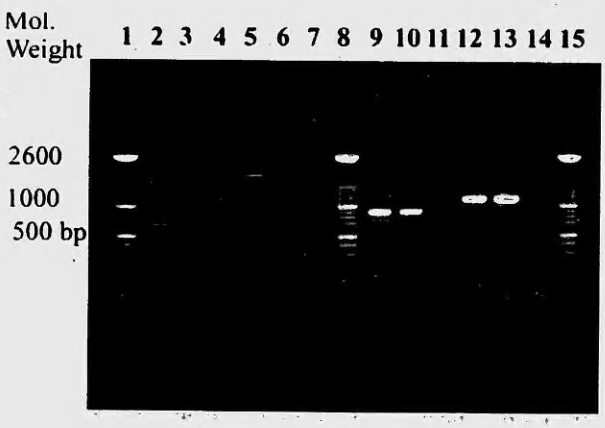


Fig. 3. Genomic DNA amplifications of the dystrophin muscle promoter and exon 1 (628 bp), dystrophin exon 79 (2100 bp), tissue inhibitor of metaloproteinase-1 (TIMP-1) intron 2 (851 bp), and retinitis pigmentosa GTP-ase regulator (RPGR) intron 9 (1150 bp). Lanes 1, 8 and 15 are molecular weight markers (100 bp ladder, Boerhinger Mannheim). Lanes 2 and 3 are dystrophin promoter and exon 1 amplifications from normal dog (positive) and dystrophic GSHP DNA (negative), respectively. Lanes 5 and 6 are dystrophin exon 79 amplifications from normal dog (positive) and affected GSHP DNA (negative), respectively. Lanes 9 and 10 are TIMP-1 intron 2 amplifications from normal dog (positive) and affected GSHP DNA (positive), respectively. Lanes 12 and 13 are RPGR intron 9 amplifications from normal dog (positive) and affected GSHP DNA (positive), respectively. Lanes 4, 7, 11 and 14 are negative controls with no template added.

3.2. Western blot analysis

Western blot analysis of a muscle biopsy taken from the triceps muscle of GSHP dog II at 5 months of age revealed a complete absence of dystrophin, whereas normal dog skeletal muscle revealed the expected 427 kDa dystrophin protein (Fig. 1). Both GSHP and control skeletal muscles revealed a 43 kDa β-dystroglycan band, although the levels were slightly reduced in the GSHP (Fig. 1). The blot labeled with the utrophin antibody showed faint 400 kDa bands in normal dog muscle but clear bands of apparently identical intensity in GSHP dog II and in all five GRMD dogs (Fig. 2).

3.3. PCR analysis

Using a canine dystrophin promoter screening strategy, we have reported previously that the first exons associated with the cortical, muscle and Purkinje dystrophin promoters could not be amplified from the GSHP genomic DNA [22]. We have now shown that neither the canine dystrophin M promoter nor dystrophin exon 79 could be amplified by PCR from either of the GSHPs (Fig. 3). Further, genomic amplifications for canine dystrophin exons 1, 7–9 and 21 were all negative (data not shown). Positive amplifications for the TIMP-1 and RPGR gene confirmed the integrity of the GSHP genomic DNA (Fig. 3). These positive amplifications also helped to establish a proximal X-chromosomal deletion border, based on the strong synteny between the canine and human X-chromosomes [26].

3.4. Cytogenetical findings

Analysis of DAPI banded metaphases from the dam revealed that one of her X-chromosomes did appear consistently shorter than the apparently normal homologue (data not shown). This reduction in size appeared to be the consequence of a reduction in the width of the G-dark band designated as Xp21 [27], although this is not sufficiently reliable to be used for confident diagnosis even in elongated chromosomes. Since the human dystrophin locus is located

in the region Xp21 and there is a high level of conservation of the mammalian X chromosome, it was decided to further investigate the presence of the deletion by FISH using canine probes, known to map in the vicinity of the deleted region.

From a panel of approximately 30 physically mapped canine microsatellite clones, isolated from a flow sorted canine X chromosome library, four clones were selected; two from within the Xp21 region (x71 and x74) and two (x9 and x36) proximal and distal to Xp21, respectively (Dickens et al., in preparation). When examined against the Genbank database using the BLAST tool, the sequence of one of these clones (x74) showed a high level of sequence homology ($4.7 \times e^{-31}$) to exons 7–9 of a human dystrophin sequence (accession number U60822). This clone and a second, anonymous, microsatellite containing clone (x71), also mapping to Xp21, could not be amplified by PCR from either of the affected GSHPs, and produced only a weak amplification in the carrier dam (data not shown). This suggested that both loci mapped to the deleted region.

FISH of all four Xp clones onto metaphase chromosome preparations of the carrier dam and her two affected male offspring, revealed that while one of the X chromosomes of the dam showed hybridization of all four clones, the other X chromosome of the dam and also the X chromosome in both male offspring showed no signal from either Xp21 clone, but did show signal from the flanking clones (Fig. 4). Further, in the non-deleted X, the two Xp21 clones were clearly separable on mid-metaphase preparations, with the dystrophin locus (x74) proximal to the anonymous STR clone (x71), suggesting that these loci are separated by at least 1.5–2 Mb.

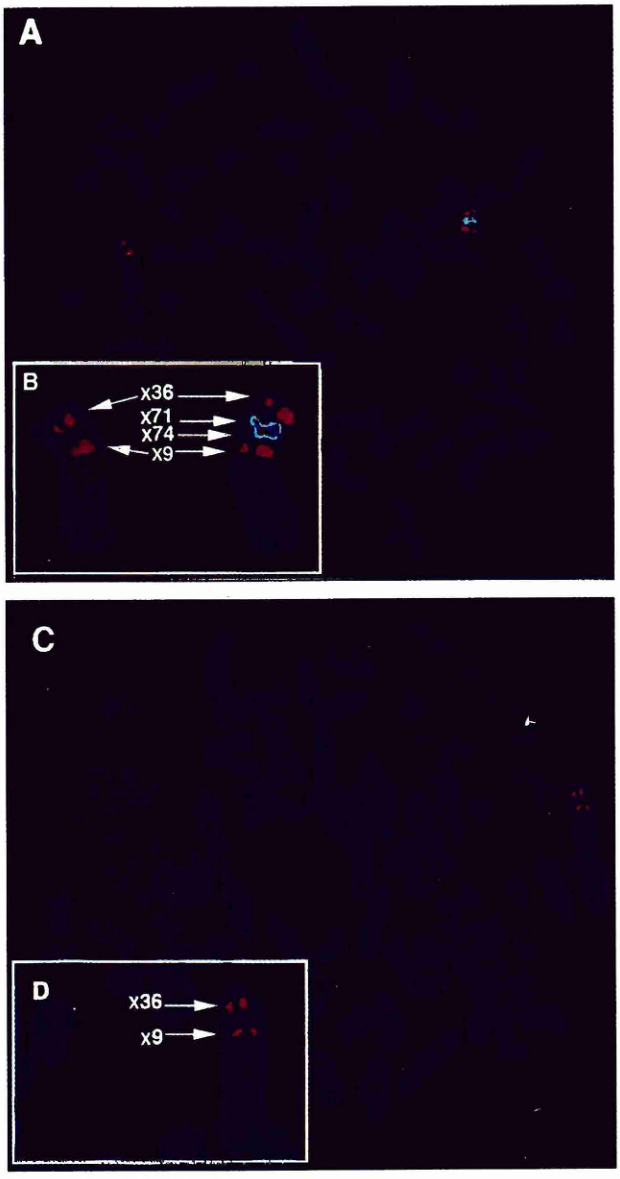


Fig. 4. FISH analysis of the carrier dam (A,B) and the affected male offspring (C,D) with Xp specific plasmid clones. Two clones (x71 and x74) both mapping to Xp21 were labeled with digoxygenin and detected with FITC (green signal) and two clones proximal (x9) and distal (x36) to Xp21 were labeled with biotin and detected with Texas Red (red signal). In (A,B) the normal X (right) of the carrier dam clearly demonstrates the presence of signal (as RED-GREEN-GREEN-RED) from the hybridization of all four clones, whereas the deleted X (left) only has signal (as RED-RED) resulting from the hybridization of the two flanking clones. The pattern of hybridization seen on the deleted X (left) of the carrier dam in (A) is also seen in the X chromosome of the male offspring (dog I) shown in (C,D). Panels B and D are enlargements of the labeled X chromosomes in panels A and C, respectively.

4. Discussion

We have identified a new canine model for DMD in two male GSHP littermates. The molecular basis for canine models of DMD has been described previously, only for the golden retriever and rottweiler mutations. Golden retriever muscular dystrophy is caused by a splice-site mutation in the sixth intron of the dystrophin gene, which alters the translational reading frame and predicts a truncated non-functional dystrophin protein product [17]. Muscular dystrophy in rottweilers is caused by an exon 58 substitution mutation that creates a premature stop codon and a truncated dystrophin [18]. We now report that dystrophic GSHPs have a novel mutation encompassing the entire dystrophin gene.

The GSHP deletion includes the entire dystrophin promoter region through the 3′ untranslated region in the 79th exon. The only other natural dystrophin deletion described in an animal model is in the dystrophic cat, which is missing approximately 200 kb of the dystrophin gene, including the M and P promoters and their respective first exons [16]. The GSHP deletion is much larger than this and spans at least 2.4 megabases of the canine X chromosome. Only a few X-chromosomal deletions of this magnitude have been identified, even in DMD patients [9–12].

Most DMD patients with extensive X-chromosomal deletions exhibit several clinical diseases that may include the McLeod syndrome, chronic granulomatous disease, glycerol kinase deficiency, adrenal insufficiency, or retinitis pigmentosa-3 (RP-3) [9–12]. The canine RPGR gene

(Xp21.1 in humans) is probably located immediately centromeric to the dystrophin gene (Xp21.2–Xp21.3), based on the strong synteny between large regions of the canine and human X-chromosomes [26]. Positive X-chromosomal amplification for the canine RPGR gene in intron 4 (and intron 9, data not shown) confirms the presence of this gene. Further, the lack of an RP-3 phenotype and our inability to identify any of the above X-linked conditions in the GSHPs suggests that the adjacent X-chromosomal disease genes are also intact.

We used FISH analysis to confirm the approximate size and extent of the X-chromosomal deletion in the dam and the affected GSHPs. The resolving power of FISH in metaphase chromosomes is roughly 1-2 Mb [28]. Data presented here suggests that the two markers contained within the deleted region of the affected X chromosome (x71 and x74) are separated by at least 2.0 Mb. The carrier GSHP dam is heterozygous for this deletion, and FISH has also unambiguously demonstrated that the deleted X chromosome was transmitted to both of her affected male pups. These data demonstrate the tremendous utility of FISH technology for the detection of sub-microscopic chromosomal deletions. Measurements of X chromosomes in DAPI banded elongated metaphases from the carrier dam revealed that the deleted X chromosome was consistently 2% smaller than its homologue. The size of the canine X chromosome has been estimated to be 137 Mb [29]; therefore, the size difference represents a loss of approximately 2.7 Mb.

This new canine deletion leads to a complete absence of dystrophin protein as observed on both immunohistochemical and Western blot analysis of affected GSHP skeletal muscle. Analysis of skeletal muscle from most DMD patients similarly reveals a complete absence of dystrophin [19,20]. In contrast, analysis of skeletal muscle from some DMD patients and from GRMD dogs reveals low levels of a near full length dystrophin species on Western blot and occasional dystrophin positive fibers on IHC [30–32]. Production of such dystrophin species in the presence of a frameshifting mutation may be explained by somatic reversion, by alternative mRNA processing, by translational reinitiation, or by a combination thereof [30,33–39]. The GSHPs cannot utilize any of these mechanisms of frameshift correction to produce a protein, as confirmed by their complete lack of dystrophin immunoreactivity.

This spontaneous canine dystrophin 'knockout' should prove useful for dystrophin gene therapy trials, myoblast transfer, or a combination of the two. Current animal models used for testing such therapies include the *mdx* mouse and GRMD dogs [40–42]. Both produce native dystrophin transcripts and protein that may complicate the interpretation of experimental results. The dystrophic GSHPs do not produce cortical (C), muscle (M), or Purkinje (P) dystrophin transcripts [22], nor can they produce any of the internal dystrophin isoforms including Dp260, Dp140, Dp116 and Dp71 [43–46]. Any dystrophin transcripts or protein detected in GSHP skeletal muscle after therapeutic intervention, could therefore only be produced by the dystrophin delivery vehicle. Furthermore, the efficacy of any trial could easily be assessed using RT PCR, immunohistochemistry, and Western blot analysis.

Despite a complete absence of dystrophin immunoreactivity in GSHP II, this dog manifests an extremely mild skeletal myopathy compared with his littermate. In light of recent studies which suggest that utrophin may ameliorate the *mdx* phenotype [47,48], we have conducted utrophin Western blot analysis to determine whether or not utrophin over-expression is responsible for rescuing the mildly affected GSHP. Unfortunately, it was not possible to obtain a muscle biopsy from GSHP dog I. We therefore conducted Western blot analysis on five GRMD dogs of varying clinical severity, including three severely affected dogs and two mildly affected dogs that were phenotypically similar to GSHP II. Utrophin is up-regulated in the GSHP skeletal muscle, as is also observed in the skeletal muscle of GRMD dogs, DMD patients [21] and the *mdx* mouse [49]. However, the utrophin blot showed clear bands of apparently identical intensity in GSHP II and all five GRMD dogs. These data do not support a role of utrophin in ameliorating the dystrophic canine phenotype. Further investigation of both canine models may provide additional insight into the utrophin rescue phenomenon.

Acknowledgements

We are grateful to Anna Bristow (A.H.T.) for her excellent technical assistance. M.B., H.F.D. and M.M.B. are supported by funds from the Guide Dogs for the Blind Association. C.F.L. is supported by the Wellcome Trust.

References

[1] Monaco A, Walker A, Millwood I, Larin Z, Lehrach H. A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene. Genomics 1992;12:465–473.
[2] Boyce F, Beggs A, Feener C, Kunkel L. Dystrophin is transcribed in brain from a distant upstream promoter. Proc Natl Acad Sci USA 1991;88:1276–1280.
[3] Gorecki D, Monaco A, Derry J, Walker A, Barnard E, Barnard P. Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. Hum Mol Genet 1992;1:505–510.
[4] Klamut H, Gangopadhyay S, Worton R, Ray P. Molecular and functional analysis of the muscle-specific promoter region of the Duchenne muscular dystrophy gene. Mol Cell Biol 1990;10:193–205.
[5] den Dunnen J, Grootscholten P, Bakker E, et al. Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. Am J Hum Genet 1989;45:835–847.
[6] Kunkel L. Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. Nature 1986;322:73–77.
[7] Beggs A, Koenig M, Boyce F, Kunkel L. Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. Hum Genet 1990;86:45–48.
[8] Chamberlain J, Gibbs R, Ranier J, Nguyen P, Caskey C. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res 1988;16:11141–11156.

[9] Francke U, Ochs H, Giacalone J, et al. Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome. Am J Hum Genet 1985;37:250–267.
[10] Clarke A, Roberts S, Thomas N, Whitfield A, Williams J, Harper P. Duchenne muscular dystrophy with adrenal insufficiency and glycerol kinase deficiency: high resolution cytogenetic analysis with molecular, biochemical, and clinical studies. J Med Genet 1986;23: 501–508.
[11] Bartley J, Patil S, Davenport S, Goldstein D, Pickens J. Duchenne muscular dystrophy, glycerol kinase deficiency, and adrenal insufficiency associated with Xp21 interstitial deletion. J Pediatr 1986;108: 189–192.
[12] Stuhrmann M, Heilbronner H, Reis A, Wegner R, Fischer P, Schmidtke J. Characterization of a Xp21 microdeletion syndrome in a 2-year-old boy with muscular dystrophy, glycerol kinase deficiency and adrenal hypoplasia congenita. Hum Genet 1991;86:414–415.
[13] Calvano S, Memeo E, Piemontese M, et al. Detection of dystrophin deletion carriers using FISH analysis. Clin Genet 1997;52:17–22.
[14] Voskova-Goldman A, Peier A, Caskey C, Richards C, Shaffer L. DMD-specific FISH probes are diagnostically useful in the detection of female carriers of DMD gene deletions. Neurology 1997;48: 1633–1638.
[15] Sicinski P, Geng Y, Ryder-Cook A, Barnard E, Darlison M, Barnard P. The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science 1989;244:1578–1580.
[16] Winand N, Edwards M, Pradhan D, Berian C, Cooper B. Deletion of the dystrophin muscle promoter in feline muscular dystrophy. Neuromuscul Disord 1994;4:433–445.
[17] Sharp N, Kornegay J, Van Camp S, et al. An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics 1992;13: 115–121.
[18] Winand N, Pradham D, Cooper B. 'Molecular characterization of severe Duchenne-type muscular dystrophy in a family of rottweiler dogs' presented at Molecular mechanisms of neuromuscular disease, Tuscon, AZ, 1994.
[19] Nicholson L, Davison K, Johnson M, et al. Dystrophin in skeletal muscle. II. Immunoreactivity in patients with Xp21 muscular dystrophy. J Neurol Sci 1989;94:137–146.
[20] Nicholson L, Davison K, Falkous G, et al. Dystrophin in skeletal muscle. I. Western Blot analysis using a monoclonal antibody. J Neurol Sci 1989;94:137–146.
[21] Man Nt, Ellis J, Love D, et al. Localization of the DMDL gene-encoded dystrophin-related protein using a panel of nineteen monoclonal antibodies: presence at neuromuscular junctions, in the sarcolemma of dystrophic skeletal muscle, in vascular and other smooth muscles, and in proliferating brain cell lines. J Cell Biol 1991;115:1695–1700.
[22] Schatzberg S, Olby N, Steingold S, et al. A polymerase chain reaction (PCR) screening strategy for the canine dystrophin promoter Am J Vet Res (in press).
[23] Zeiss C, Acland G, Aguirre G, Ray K. Cloning of RPGR in a canine model of XLRP and characterization of novel splice variants. Invest Ophthalmol Vis Sci 1998;39:4082.
[24] Breen M, Arveiler B, Murray I, Gosden J, Porteous D. YAC mapping by FISH using Alu-PCR-generated probes. Genomics 1992;13:726–730.
[25] Moise N, Valentine B, Brown C, et al. Duchenne's cardiomyopathy in a canine model: electrocardiographic and echocardiographic studies. J Am Coll Cardiol 1991;17:812–820.
[26] Deschenes S, Puck J, Dutra A, Somberg R, Felsburg P, Henthorn P. Comparative mapping of canine and human proximal Xq and genetic analysis of canine X-linked severe combined immunodeficiency. Genomics 1994;23:62–68.
[27] Reimann N, Bartnitzke S, Bullerdiek J, et al. An extended nomenclature of the canine karyotype. Cytogenet Cell Genet 1996;73:140–144.
[28] Giovannini M, Selleri L, Biegel J, Scotlandi K, Emanuel B, Evans G. Interphase cytogenetics for the detection of the t(11;22) (q24;q12) in small round cell tumors. J Clin Invest 1992;90:1911–1918.
[29] Langford C, Fischer P, Binns M, Holmes N, Carter N. Chromosome-specific paints from a high-resolution flow karyotype of the dog. Chromosome Res 1996;4:115–123.
[30] Schatzberg S, Anderson L, Wilton S, et al. Alternative dystrophin gene transcripts in golden retriever muscular dystrophy. Muscle Nerve 1998;21:991–998.
[31] Nicholson L. The 'rescue' of dystrophin synthesis in boys with Duchenne muscular dystrophy. Neuromuscul Disord 1993;3:525–531.
[32] Kornegay J, Sharp N, Bogan D, Van Camp S, Metcalf J, Schuler R. Contraction tension and kinetics of the peroneus longus muscle in golden retriever muscular dystrophy. J Neurol Sci 1994;123:100–107.
[33] Winnard A, Mendell J, Prior T, Florence J, Burghes A. Frameshift deletions of exons 3–7 and revertant fibers in Duchenne muscular dystrophy: mechanisms of dystrophin production. Am J Hum Genet 1995;56:158–166.
[34] Sherratt T, Vulliamy T, Dubowitz V, Sewry C, Strong P. Exon skipping and translation in patients with frameshift deletions in the dystrophin gene. Am J Hum Genet 1993;53:1007–1015.
[35] Wilton S, Dye D, Laing N. Dystrophin gene transcripts skipping the mdx mutation. Muscle Nerve 1997;20:728–734.
[36] Hoffman E, Morgan J, Watkins S, Partridge T. Somatic reversion/suppression of the mouse mdx phenotype in vivo. J Neurol Sci 1990; 99:9–25.
[37] Chelly J, Gilgenkrantz H, Lambert M, et al. Effect of dystrophin gene deletions on mRNA levels and processing in Duchenne and Becker muscular dystrophies. Cell 1990;63:1239–1248.
[38] Thanh LT, Man Nt, Helliwell T, Morris G. Characterization of revertant muscle fibers in Duchenne muscular dystrophy, using exon-specific monoclonal antibodies against dystrophin. Am J Hum Genet 1995;56:725–731.
[39] Fanin M, Danieli G, Vitiello L, Senter L, Angelini C. Prevalence of dystrophin-positive fibers in 85 Duchenne muscular dystrophy patients. Neuromuscul Disord 1992;2:41–45.
[40] Gussoni E, Blau H, Kunkel L. The fate of individual myoblasts after transplantation into muscles of DMD patients. Nat Med 1997;3:970–977.
[41] Howell J, Fletcher S, Kakulas B, O'Hara M, Lochmuller H, Karpati G. Use of the dog model for Duchenne muscular dystrophy in gene therapy trials. Neuromuscul Disord 1997;7:325–328.
[42] Fassati A, Wells D, Sgro SP, et al. Genetic correction of dystrophin deficiency and skeletal muscle remodeling in adult MDX mouse via transplantation of retroviral producer cells. J Clin Invest 1997;100:620–628.
[43] Rapaport D, Lederfein D, Grootscholten P, et al. Characterization and cell type distribution of a novel, major transcript of the Duchenne muscular dystrophy gene. Differentiation 1992;49:187–193.
[44] D'Souza V, Nguyen T, Morris G, Karges W, Pillers D, Ray P. A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet 1995;4:837–842.
[45] Byers, T., Lidov, H., Kunkel, L. An alternative dystrophin transcript specific to peripheral nerve. Nat Genet 1993;4:1,77–82.
[46] Lidov H, Selig S, Kunkel L. Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. Hum Mol Genet 1995;4:329–335.
[47] Deconinck N, Tinsley J, Fisher R, et al. Expression of truncated utrophin leads to major functional improvements in dystrophin-deficient muscles of mice. Nat Med 1997;3:1216–1221.
[48] Tinsley J, Potter A, Phelps S, Fisher R, Trickett J, Davies K. Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene (see comments). Nature 1996;384:349–353.
[49] Pons F, Robert A, Marini J, Leger J. Does utrophin expression in muscles of mdx mice during postnatal development functionally compensate for dystrophin deficiency? J Neurol Sci 1994;122: 162–170.